

Probabilistic existence of regular combinatorial structures*

Greg Kuperberg[†]

Shachar Lovett[‡]

Ron Peled[§]

February 19, 2013

Abstract

We show the existence of regular combinatorial objects which previously were not known to exist. Specifically, for a wide range of the underlying parameters, we show the existence of non-trivial orthogonal arrays, t -designs, and t -wise permutations. In all cases, the sizes of the objects are optimal up to polynomial overhead. The proof of existence is probabilistic. We show that a randomly chosen structure has the required properties with positive yet tiny probability. Our method allows also to give rather precise estimates on the number of objects of a given size and this is applied to count the number of orthogonal arrays, t -designs and regular hypergraphs. The main technical ingredient is a special local central limit theorem for suitable lattice random walks with finitely many steps.

*An extended abstract [KLP12] of this paper appeared in the 44th ACM Symposium on Theory of Computing (STOC 2012).

[†]University of California, Davis. E-mail: greg@math.ucdavis.edu. Supported by NSF grant CCF-1013079.

[‡]Institute for Advanced Study. E-mail: slovett@math.ias.edu. Supported by NSF grant DMS-0835373.

[§]Tel Aviv University, Israel. E-mail: peledron@post.tau.ac.il. Supported by an ISF grant and an IRG grant.

Contents

1	Introduction	3
1.1	Orthogonal arrays	4
1.2	Designs	5
1.2.1	Regular hypergraphs	6
1.3	Permutations	7
1.4	Proof overview	7
1.5	Related work	9
1.6	Paper organization	9
2	General framework	10
3	Applications	12
3.1	Local decodability	13
3.2	Orthogonal arrays	13
3.3	Designs	17
3.4	t -wise permutations	21
3.4.1	Irreducible representations of the symmetric group	22
3.4.2	Two bases for W and the divisibility constant	23
3.4.3	Antisymmetrizers	25
3.4.4	Spanning vectors for W^\perp	26
3.5	The number of t -wise permutations	27
4	Proof of main theorems	28
4.1	Local central limit theorem statement	28
4.2	Fourier analysis	29
4.3	Local correction	31
4.4	Estimating the Fourier transform near zero	33
4.5	Bounding the Fourier transform far from L	34
4.6	Proof of central limit theorem from auxiliary lemmas	35
4.7	Proof of main theorems	39
4.8	Basis-free formulation of local central limit theorem	40
5	Summary and open problems	41

1 Introduction

We introduce a new framework for establishing the existence of regular combinatorial structures, such as orthogonal arrays, t -designs and t -wise permutations. Let B be a finite set and let V be a vector space of functions from B to the rational numbers \mathbb{Q} . We study when there is a small subset $T \subset B$ satisfying

$$\frac{1}{|T|} \sum_{t \in T} f(t) = \frac{1}{|B|} \sum_{b \in B} f(b) \quad \text{for all } f \text{ in } V. \quad (1)$$

In probabilistic terminology, equation (1) means that if t is a uniformly random element in T and b is a uniformly random element in B then

$$\mathbb{E}_{t \in T}[f(t)] = \mathbb{E}_{b \in B}[f(b)] \quad \text{for all } f \text{ in } V, \quad (2)$$

where \mathbb{E} denotes expectation. Of course, (1) holds trivially when $T = B$. Our goal is to find conditions on B and V that yield a small subset T that satisfies (1), where in our situations, small will mean polynomial in the dimension of V . We remark that in many natural problems one might encounter a vector space V over \mathbb{R} or \mathbb{C} instead. However, since (1) is a rational equation, we can always reduce to the case of rational vector spaces.

A more concrete realization of the above framework is given by the following problem. Let ϕ be a matrix with rational entries whose rows are indexed by a finite set B and whose columns are indexed by a finite set A . When is there a small subset $T \subset B$ such that the average of the rows indexed by T equals the average of all rows? This problem is a special case of the above framework with V being the subspace spanned by the columns of ϕ . In fact, the general framework can always be reduced to such a problem by choosing a basis (ϕ_a) , $a \in A$, of V and defining the matrix ϕ by $\phi(b, a) = \phi_a(b)$.

Our main theorem, Theorem 2.4, gives sufficient conditions for the existence of a small subset T satisfying (1). A second theorem, Theorem 2.5, provides sharp estimates on the number of such subsets of a given size. We apply the theorems to establish results in three interesting cases of the general framework: orthogonal arrays, t -designs, and t -wise permutations. These are defined and discussed in more detail in the next sections. Our methods solve an open problem, whether there exist non-trivial t -wise permutations for every t . They strengthen Teirlinck's theorem [Tei87], which was the first theorem to show the existence of t -designs for every t . And they improve existence results for orthogonal arrays, when the size of the alphabet is divisible by many distinct primes. Moreover, in all three cases considered, we show the existence of a structure whose size is optimal up to polynomial overhead. In addition, we provide sharp estimates for the number of orthogonal arrays and t -designs of a given size. As a special case, these yield estimates for the number of regular hypergraphs of a given degree.

Our approach to the problem is via probabilistic arguments. In essence, we prove that a random subset of B satisfies equation (1) with positive, albeit tiny, probability. Thus our method is one of the few known methods for showing existence of rare objects. This class includes such other methods as the Lovász local lemma [EL75] and Spencer's "six deviations suffice" method [Spe85]. However, our method does not rely on these previous approaches. Instead, our technical ingredient is a special version of the (multi-dimensional) local central limit theorem with finitely many steps. We cannot use any "off the shelf" local central limit theorem, not even one enhanced by a Berry-Esseen-type estimate of the rate of convergence, since the number of steps of our random walk is small compared to the dimension of the space in which it takes its values. Instead, we prove the local central limit theorem that we need directly using Fourier analysis. Section 1.4 gives an overview of our approach.

We also mention that efficient randomized algorithm versions of the Lovász local lemma [Mos09, MT10] and Spencer's method [Ban10, LM12] have recently been found. Relative to these new algorithms, the objects that they produce are no longer rare. Our method is the only one that we know that shows the existence of rare combinatorial structures, which are still rare relative to any known, efficient, randomized algorithm.

1.1 Orthogonal arrays

Here and in the rest of the paper we use the notation $[m] := \{1, \dots, m\}$. A subset $T \subset [q]^n$ is an *orthogonal array of alphabet size q , length n and strength t* if it yields all strings of length t with equal frequency if restricted to any t coordinates. In other words, for any distinct indices $i_1, \dots, i_t \in [n]$ and any (not necessarily distinct) values $v_1, \dots, v_t \in [q]$,

$$|\{x \in T : x_{i_1} = v_1, \dots, x_{i_t} = v_t\}| = q^{-t}|T|. \quad (3)$$

Equivalently, choosing $x = (x_1, \dots, x_n) \in T$ uniformly, the distribution of each coordinate of x is uniform in $[q]$ and every t coordinates of x are independent (x is t -wise independent). For an introduction to orthogonal arrays see [HSS99].

Orthogonal arrays fit into our general framework as follows. We take B to be $[q]^n$ and V to be the space spanned by all functions of the form

$$f_{(I,v)}(x_1, \dots, x_n) = \begin{cases} 1 & x_i = v_i \text{ for all } i \in I \\ 0 & \text{Otherwise} \end{cases},$$

with $I \subset [n]$ a subset of size t and $v \in [q]^I$. With this choice, a subset $T \subset B$ satisfying (1) is precisely an orthogonal array of alphabet size q , length n and strength t .

It is well known that if $T \subset [q]^n$ is an orthogonal array of strength t then $|T| \geq (\frac{cqn}{t})^{t/2}$ for some universal constant $c > 0$ (see, e.g., [Rao73]). Matching constructions of size $|T| \leq q^{ct} (\frac{n}{t})^{c_q t}$ are known, however, as these rely on finite field properties, the constant c_q generally tends to infinity with the number of distinct prime factors of q . Our technique provides the first upper bound on the size of orthogonal arrays in which the constant in the exponent is independent of q . Here and below, a universal constant is a constant independent of all other parameters.

Theorem 1.1 (Existence of orthogonal arrays). *For all integers $q \geq 2$, $n \geq 1$ and $1 \leq t \leq n$ there exists an orthogonal array T of alphabet size q , length n and strength t satisfying $|T| \leq (\frac{cqn}{t})^{ct}$ for some universal constant $c > 0$.*

Moreover, we provide a rather precise count of the number of orthogonal arrays with given parameters. Such a count appears to be new even in the most well-studied case when $q = 2$.

Theorem 1.2 (Number of orthogonal arrays). *There exists a constant $c > 0$ such that for all integers $q \geq 2$, $n \geq 1$ and $1 \leq t \leq n$ and for all N satisfying that N is a multiple of q^t and $\min(N, q^n - N) \geq (\frac{cqn}{t})^{ct}$, we have that the number of orthogonal arrays T of alphabet size q , length n , strength t and $|T| = N$ equals*

$$\frac{q^{-\frac{1}{2}n \binom{n-1}{t} (q-1)^t}}{(2\pi p(1-p))^{\frac{1}{2} \sum_{i=0}^t \binom{n}{i} (q-1)^i} p^N (1-p)^{q^n - N}} (1 + \delta)$$

where $p := \frac{N}{q^n}$ and $|\delta| \leq \frac{(\frac{cqn}{t})^{ct}}{\sqrt{\min(N, q^n - N)}}$.

As a final remark we note that any orthogonal array T of alphabet size q and strength t must satisfy that $|T|$ is a multiple of q^t by (3). Hence the essential restriction on N in the last theorem is only that N be bounded away from 0 and q^n .

1.2 Designs

A (simple) t -(v, k, λ) *design*, or t -design for short, is a family of distinct subsets of $[v]$, where each set is of size k , such that each t elements belong to exactly λ sets. In other words, denoting by $\binom{[v]}{k}$ the family of all subsets of $[v]$ of size k , a set $T \subset \binom{[v]}{k}$ is a t -design if for any distinct elements $i_1, \dots, i_t \in [v]$,

$$|\{s \in T : i_1, \dots, i_t \in s\}| = \lambda. \quad (4)$$

It follows that λ satisfies the relation

$$\lambda = \frac{\binom{k}{t}}{\binom{v}{t}} |T|. \quad (5)$$

For an introduction to combinatorial designs see [CD07].

Our general framework includes t -designs as follows. We take B to be $\binom{[v]}{k}$ and V to be the space spanned by all functions of the form

$$f_a(b) = \begin{cases} 1 & a \subset b \\ 0 & \text{Otherwise} \end{cases},$$

with $a \in \binom{[v]}{t}$. With this choice, a subset $T \subset B$ satisfying (1) is precisely a simple t -(v, k, λ) design, with λ given by (5).

Although t -designs have been investigated for many years, the basic question of existence of a design for a given set of parameters t, v, k and λ remains mostly unanswered unless t is quite small. The case $t = 2$ is known as a block design and much more is known about it than for larger t . Explicit constructions of t -designs for $t \geq 3$ are known for various specific constant settings of the parameters (e.g. 5-(12, 6, 1) design). The breakthrough result of Teirlinck [Tei87] was the first to establish the existence of non-trivial t -designs for $t \geq 7$. In Teirlinck's construction, $k = t + 1$ and v satisfies congruences that grow very quickly as a function of t . Other sporadic and infinite examples have been found since then (see [CD07] or [Mag09] and the references within), however, the set of parameters which they cover is still very sparse.

It follows from (5) that any t -(v, k, λ) design T has size $|T| = \lambda \binom{v}{t} / \binom{k}{t} \geq \left(\frac{v}{k}\right)^t$. Moreover, it can be shown [RCW75] that whenever $v \geq k + t$ the inequality $|T| \geq \binom{v}{\lfloor t/2 \rfloor} \geq \left(\frac{v}{t}\right)^{\lfloor t/2 \rfloor}$ holds. Even when existence has been shown, the designs obtained are often inefficient in the sense that their size is much larger than these lower bounds permit. One of the main results of our work is to establish the existence of efficient t -designs for a wide range of parameters.

Theorem 1.3 (Existence of t -designs). *For all integers $v \geq 1$, $1 \leq t \leq v$ and $t \leq k \leq v$ there exists a t -(v, k, λ) design whose size is at most $\left(\frac{cv}{t}\right)^{ct}$ for some universal constant $c > 0$.*

Our work also provides a rather precise count of the number of t -designs of a given size with given parameters. To state this count precisely, we recall the well-known observation that if T is a t -(v, k, λ) design then for every $1 \leq s \leq t$, each subset of size s in $[v]$ is covered by exactly

$$\lambda_s := \frac{\binom{k}{s}}{\binom{v}{s}} |T|$$

sets in T . In particular, (λ_s) , $1 \leq s \leq t$ must be integers. Our next theorem, in addition to estimating the number of designs, implies that if $|T|$ (or equivalently λ) is sufficiently large, these integrality conditions suffice for the existence of t -(v, k, λ) designs.

Theorem 1.4 (Number of t -designs). *There exists a constant $c > 0$ such that for all integers $v \geq 1$, $1 \leq t \leq v$ and $t \leq k \leq v$ and for all N satisfying that the numbers*

$$\frac{\binom{k}{s}}{\binom{v}{s}} N \text{ are integers for } 1 \leq s \leq t$$

and satisfying that $\min(N, \binom{v}{k} - N) \geq \left(\frac{cv}{t}\right)^{ct}$, we have that the number of t -(v, k, λ) designs T of size $|T| = N$ equals

$$\frac{1}{(2\pi p(1-p))^{\frac{1}{2}\binom{v}{t}} p^N (1-p)^{\binom{v}{k}-N}} \left(\prod_{s=0}^t \left[\frac{\binom{k-s}{k-t}}{\binom{v-t-s}{k-t}} \right]^{\frac{1}{2}(\binom{v}{s}-\binom{v}{s-1})} \right) (1+\delta)$$

where $p := \frac{N}{\binom{v}{k}}$, $|\delta| \leq \frac{\left(\frac{cv}{t}\right)^{ct}}{\sqrt{\min(N, \binom{v}{k}-N)}}$ and $\binom{v}{-1}$ is defined to be 0.

1.2.1 Regular hypergraphs

A (simple) k -uniform hypergraph on n vertices is a family of sets of size k (called edges) on n elements (called vertices). A hypergraph is d -regular if each vertex belongs to exactly d edges. It is straightforward to check that d -regular, k -uniform hypergraphs are the same as 1 -(n, k, d) designs.

The existence question for d -regular, k -uniform hypergraphs is quite simple, such a hypergraph exists if and only if nd is divisible by k , and in this case any such hypergraph has exactly $\frac{nd}{k}$ edges. However, counting the number of d -regular, k -uniform hypergraphs is a non-trivial problem which has received much attention in the literature, mainly in the graph case ($k=2$). See, e.g., the paper by Wormald and McKay [MW90] and references within. In the graph case, approximate counts are known when the graphs are either somewhat sparse, or very dense.

Since d -regular, k -uniform hypergraphs are a special case of t -designs, we may translate Theorem 1.4 to obtain a count of such hypergraphs. Our result applies for k and d sufficiently large and appears to be new.

Theorem 1.5. *There exists a constant $c > 0$ such that for all integers $n \geq 2$, $1 \leq k \leq n$ and $1 \leq d \leq \binom{n-1}{k-1}$ satisfying that nd is divisible by k and $\min(\frac{nd}{k}, \binom{n}{k} - \frac{nd}{k}) \geq n^c$, we have that the number of d -regular, k -uniform hypergraphs on n vertices equals*

$$\frac{1}{(2\pi p(1-p))^{\frac{n}{2}} p^{\frac{nd}{k}} (1-p)^{\binom{n}{k}-\frac{nd}{k}}} \left(\frac{k}{\binom{n-1}{k-1}} \right)^{\frac{1}{2}} \left(\frac{1}{\binom{n-2}{k-1}} \right)^{\frac{1}{2}(n-1)} (1+\delta)$$

where $p := \frac{nd}{k\binom{n}{k}}$ and $|\delta| \leq \frac{n^c}{\sqrt{\min(\frac{nd}{k}, \binom{n}{k}-\frac{nd}{k})}}$.

We mention also a related result on the asymptotic formula for the number of *binary contingency tables* obtained in [CM05]. A binary contingency table is an $N \times n$ matrix with entries in $\{0, 1\}$, all row sums equal to k and all column sums equal to d (so that necessarily $N = \frac{nd}{k}$). Such a matrix describes a d -regular, k -uniform hypergraph on n vertices with *labelled* edges and allowing multiple edges. Equivalently, it describes a bipartite graph with N vertices of degree k on one side and n

vertices of degree d on the other side. When the parameters (N, n, k, d) are such that most binary contingency tables have all rows distinct, the number of such tables is close to $N!$ times the number of d -regular, k -uniform hypergraphs on n vertices. Thus our result can probably be used to extend the asymptotic formula of [CM05] to additional cases. We do not develop this direction here.

1.3 Permutations

A family of permutations $T \subset S_n$ is called a *t-wise permutation* if its action on any t -tuple of elements is uniform. In other words, for any distinct elements $i_1, \dots, i_t \in [n]$ and distinct elements $j_1, \dots, j_t \in [n]$,

$$|\{\pi \in T : \pi(i_1) = j_1, \dots, \pi(i_t) = j_t\}| = \frac{1}{n(n-1) \cdots (n-t+1)} |T|. \quad (6)$$

Our general framework includes t -wise permutations as follows. We take $B = S_n$ and V to be the space spanned by all functions of the form

$$f_{(i,j)}(b) = \begin{cases} 1 & b(i_1) = j_1, \dots, b(i_t) = j_t \\ 0 & \text{Otherwise} \end{cases},$$

where $i = (i_1, \dots, i_t)$ and $j = (j_1, \dots, j_t)$ are t -tuples of distinct elements in $[n]$. With this choice, a subset $T \subset B$ satisfying (1) is precisely a t -wise permutation.

Equation (6) yields a lower bound on the size of t -wise permutations, $|T| \geq n(n-1) \cdots (n-t+1)$. Constructions of families of t -wise permutations matching this lower bound are known for $t = 1, 2, 3$: the group of cyclic shifts $x \mapsto x + a$ modulo n is a 1-wise permutation; the group of invertible affine transformations $x \mapsto ax + b$ over a finite field \mathbb{F} yields a 2-wise permutation; and the group of Möbius transformations $x \mapsto (ax + b)/(cx + d)$ with $ad - bc = 1$ over the projective line $\mathbb{F} \cup \{\infty\}$ yields a 3-wise permutation. However, it is known (see, e.g., [Cam95], Theorem 5.2) that for $n \geq 25$ and $t \geq 4$ there are no *subgroups* of S_n which form a t -wise permutation (of any size), other than S_n itself and the alternating group A_n . Moreover, for $t \geq 4$ (and n large enough), no non-trivial constructions of t -wise permutations are known at all [KNR05, AL11], with the exception of the recent work [FPY12] which constructs rather large, but non-trivial, t -wise permutations of size t^{2n} for infinitely many values of n and t . One of our main results is the existence of small t -wise permutations for all n and t .

Theorem 1.6 (Existence of t -wise permutations). *For all integers $n \geq 1$ and $1 \leq t \leq n$ there exists a t -wise permutation $T \subset S_n$ satisfying $|T| \leq (cn)^{ct}$ for some universal constant $c > 0$.*

We leave the problem of estimating the number of t -wise permutations of a given size for future work. See Section 3.5 for numerical calculations related to this problem.

1.4 Proof overview

The idea of our approach is as follows. Let us first consider the following slight simplification of our main idea. Let T be a random multiset of B of some fixed size N chosen by sampling B uniformly and independently N times (with replacement). Let $(\phi_a)_{a \in A}$ be a basis of integer-valued functions for V (where A is some arbitrary finite index set). Observe that T satisfies (1) if and only if

$$\sum_{t \in T} \phi_a(t) = \frac{N}{|B|} \sum_{b \in B} \phi_a(b) = \mathbb{E} \left[\sum_{t \in T} \phi_a(t) \right] \quad \text{for all } a \text{ in } A, \quad (7)$$

where we add terms multiple times if they appear in T multiple times. Thus, defining an integer-valued random variable

$$X_a := \sum_{t \in T} \phi_a(t)$$

and $X := (X_a)_{a \in A} \in \mathbb{Z}^A$ we see that existence of a multiset of size N satisfying (1) will follow if we can show that $\mathbb{P}[X = \mathbb{E}[X]] > 0$. To this end we examine more closely the distribution of X . Let t_1, \dots, t_N be the random elements chosen in forming T . The basis $(\phi_a)_{a \in A}$ defines a mapping $\phi : B \rightarrow \mathbb{Z}^A$ by the trivial

$$\phi(b)_a := \phi_a(b).$$

Observe that our choice of random model implies that the vectors $(\phi(t_i))_{i \in [N]}$ are independent and identically distributed. Hence,

$$X = \sum_i \phi(t_i) \tag{8}$$

may be viewed as the end position of an N -step random walk in the lattice \mathbb{Z}^A . Thus we may hope that if N is sufficiently large, then X has an approximately (multi-dimensional) Gaussian distribution by the central limit theorem. If the relevant local central limit theorem holds as well, then the probability $\mathbb{P}[X = x]$ also satisfies a Gaussian approximation. In particular, since a (non-degenerate) Gaussian always has positive density at its expectation, we could conclude that $\mathbb{P}[X = \mathbb{E}[X]] > 0$ as desired. Moreover, to estimate the number of multisets of size N satisfying (1) we need only estimate $\mathbb{P}[X = \mathbb{E}[X]]$ using the Gaussian approximation.

The above description is the essence of our approach. The main obstacle is, of course, pointed out in the last step. We must control the rate of convergence of the local central limit theorem well enough so that the convergence error does not outweigh the probability density of the Gaussian distribution at $\mathbb{E}[X]$. Recall that the order of magnitude of such a density is typically $c^{-|A|}$ for some constant $c > 1$, and recall that $|A|$ is the dimension of V , which is the main parameter of our problem. So we indeed have very small probabilities. For this reason, and because we want convergence when N is only polynomial in the dimension of V , we were unable to use any standard local central limit theorem. Instead, we develop an ad hoc version using Fourier analysis.

In our proof of the main theorem, we modify the above description in one respect. It is technically more convenient to work with a slightly different probability model. Instead of choosing T as above, we set $p := N/|B|$ and define T by taking each element of B into T independently with probability p . This has the benefit of guaranteeing that T is a proper set instead of a multiset. However, it has also the disadvantage that it does not guarantee that $|T| = N$. To remedy this, we assume that the space V contains the constant function $h \equiv 1$; or if not, we can add it to V at the minor cost of increasing the dimension of V by 1. With this assumption, since

$$\mathbb{E} \left[\sum_{t \in T} h(t) \right] = \mathbb{E}[|T|] = N,$$

we see that (7), or equivalently $X = \mathbb{E}[X]$, implies both that $|T| = N$ and that (1) holds. Another disadvantage is that in this new probability model, the vector X is no longer a sum of identically distributed variables. However, since the summands in (8) are still independent, we can continue to use Fourier analysis methods in our proof.

We cannot expect there to always be a small subset $T \subset B$ that satisfies (1). For instance, Alon and Vu [AV97] found a regular hypergraph with n vertices and $\approx n^{n/2}$ edges, with no non-trivial regular sub-hypergraph. Here, a regular hypergraph is one in which every vertex belongs to the same number of hyperedges. We may describe their example in our language by letting B be the

set of hyperedges of this hypergraph, A be its vertex set, and define $\phi : B \rightarrow \{0, 1\}^A$ by letting $\phi(b)$ be the indicator function of the set of vertices contained in b . The result of [AV97] implies that while the vector $\sum_{b \in B} \phi(b)$ is constant, this property is not shared by $\sum_{t \in T} \phi(t)$ for any non-empty, proper subset $T \subset B$. Thus, we need to impose certain conditions on B and V , or equivalently on the map ϕ .

We will require certain divisibility, boundedness and symmetry assumptions. Our main theorem shows that these conditions suffice to yield the existence of a small solution of (1) and, moreover, a rather precise estimate for the number of solutions. The conditions and the statement of the theorem appear in Section 2. The existence and counting theorems for orthogonal arrays, t -designs and t -wise permutations follow by verifying these conditions for the choice of B and V detailed in Sections 1.1 through 1.3.

1.5 Related work

In the probabilistic formulation (2) of our problem we seek a small subset $T \subset B$ such that the uniform distribution over T simulates the uniform distribution over B with regards to certain tests. There are two ways to relax the problem to make its solution easier, and raise new questions regarding explicit solutions.

One relaxation is to allow a set T with a *non-uniform* distribution μ which simulates the *uniform* distribution over B . For many practical applications of t -designs and t -wise permutations in statistics and computer science, but not quite every application, this relaxation is as good as the uniform question. The existence of a solution with small support is guaranteed by Carathéodory's theorem, using the fact that the constraints on μ are all linear equalities and inequalities. Moreover, such a solution can be found efficiently, as was shown by Karp and Papadimitriou [KP82] and in more general settings by Koller and Megiddo [KM94]. Alon and Lovett [AL11] give a strongly explicit analog of this in the case of t -wise permutations and more generally in the case of group actions.

A different relaxation is to require the uniform distribution over T to only approximately satisfy equation (2). Then it is trivial that a sufficiently large random subset $T \subset B$ satisfies the requirement with high probability, and the question is to find an explicit solution. For instance, we can relax the problem of t -wise permutations to *almost* t -wise permutations. For this variant an optimal solution (up to polynomial factors) was achieved by Kaplan, Naor and Reingold [KNR05], who gave a construction of such an almost t -wise permutation of size $n^{O(t)}$.

A framework for studying problems similar to ours in a continuous setup was introduced by Seymour and Zaslavsky [SZ84]. They consider the case when B is a path-connected topological space with a measure μ of full support, and the problem of finding a finite subset $T \subset B$ such that the uniform distribution over T integrates certain continuous functions exactly as μ . This framework is sometimes referred to as “averaging sets”, “equal-weight quadratures” or “Chebyshev-type quadratures”. It was shown in [SZ84] that such T exist in great generality and that their cardinality can be any number with finitely many exceptions. However, no quantitative upper bound on $|T|$ is known in this generality.

1.6 Paper organization

We give a precise description of the general framework and our main theorem in Section 2. We apply it to show the existence and estimate the number of orthogonal arrays, t -designs and t -wise permutations in Section 3. The proof of our main theorem is given in Section 4. We summarize and give some open problems in Section 5.

2 General framework

Let B be a finite set and let V be a linear subspace of functions $f : B \rightarrow \mathbb{Q}$. The goal of this work is to find sufficient conditions for the existence of a small set $T \subset B$ such that

$$\frac{1}{|T|} \sum_{t \in T} f(t) = \frac{1}{|B|} \sum_{b \in B} f(b) \quad \text{for all } f \text{ in } V, \quad (9)$$

and moreover to estimate the number of such sets of a given size. We now describe a list of conditions on V which will be sufficient for this task. Some of these conditions are easy to verify in applications, while others require some effort. We stress that in all of our applications these properties are verified explicitly; this is in contrast with the fact that we do not know how to find T explicitly.

Divisibility. For (9) to hold for a set T with $|T| = N$ we must have

$$\sum_{t \in T} f(t) = \frac{N}{|B|} \sum_{b \in B} f(b) \quad \text{for all } f \text{ in } V.$$

In particular, we must have that

$$\text{there exists a } \gamma \in \mathbb{Z}^B \text{ such that } \sum_{b \in B} \gamma_b f(b) = \frac{N}{|B|} \sum_{b \in B} f(b) \text{ for all } f \text{ in } V. \quad (10)$$

The set of all integers N satisfying (10) consists of all integer multiples of some minimal positive integer c_1 . To see this, observe that if N_1 and N_2 are solutions then their difference is also a solution. It follows that $|T|$ must be an integer multiple of c_1 . This is the divisibility condition that we require, and we call this c_1 the *divisibility constant of V* .

We remark that to check the divisibility condition in practice it suffices to check (10) for f belonging to some basis of V . More explicitly, we make the following definition.

Definition 2.1. Let $\phi : B \rightarrow \mathbb{Q}^A$ for some finite set A . We define $\mathcal{L}(\phi)$ to be the lattice spanned by $\phi(b)$, $b \in B$. I.e.,

$$\mathcal{L}(\phi) := \left\{ \sum_{b \in B} n_b \cdot \phi(b) : n_b \in \mathbb{Z} \right\} \subset \mathbb{Q}^A.$$

Using this definition, if $\phi : B \rightarrow \mathbb{Q}^A$ is such that the vectors (ϕ_a) , $a \in A$, form a basis for V , then the divisibility condition is equivalent to having $\frac{N}{|B|} \sum_{b \in B} \phi(b) \in \mathcal{L}(\phi)$.

Boundedness. We make the following definition.

Definition 2.2. Let $W \subset \mathbb{Q}^B$ be a vector space. For $1 \leq p \leq \infty$, we say that W has a c -bounded integer basis in ℓ_p if W is spanned by integer vectors whose ℓ_p norm is at most c . That is, if

$$\text{Span}(\{\gamma \in W \cap \mathbb{Z}^B : \|\gamma\|_p \leq c\}) = W.$$

We will only use in this paper the norms $\|\gamma\|_1 = \sum_{b \in B} |\gamma_b|$ and $\|\gamma\|_\infty = \max_{b \in B} |\gamma_b|$. We denote by V^\perp the orthogonal complement of V in \mathbb{Q}^B , that is,

$$V^\perp := \{g \in \mathbb{Q}^B : \sum_{b \in B} f(b)g(b) = 0 \quad \forall f \in V\}.$$

We impose the conditions that for some small c_2 and c_3 , V has a c_2 -bounded integer basis in ℓ_∞ and V^\perp has a c_3 -bounded integer basis in ℓ_1 .

In our applications, the boundedness condition for V follows directly from the definition. However, the boundedness condition for V^\perp is less trivial and requires far more work to check. We view this condition as an analog of the LDPC (Low Density Parity Check) condition in coding theory, when viewed over the integers, and we develop techniques based on coding theory in order to guarantee it. In particular, we show in Section 3.1 that the condition is implied by a certain local decodability property of V .

Symmetry. The next condition relates to the symmetries of the subspace V .

Definition 2.3. A symmetry of V is a permutation $\pi \in S_B$ satisfying $f \circ \pi \in V$ for all $f \in V$.

Equivalently, if $\phi : B \rightarrow \mathbb{Q}^A$ is such that the vectors (ϕ_a) , $a \in A$, form a basis for V then a permutation $\pi \in S_B$ is a symmetry of V if and only if there exists an invertible linear map $\tau : \mathbb{Q}^A \rightarrow \mathbb{Q}^A$ such that

$$\phi(\pi(b)) = \tau(\phi(b)) \quad \text{for all } b \in B.$$

It is straightforward to check that the set of symmetries of V forms a subgroup of S_B . We impose the condition that this group acts transitively on B . That is, that for any $b_1, b_2 \in B$ there exists a symmetry π of V satisfying $\pi(b_1) = b_2$. In our applications this condition follows from the symmetric nature of their description.

Constant functions. Our last condition is required for somewhat technical reasons as explained in the proof overview section. We require the constant functions to belong to V . We note that this is the case in all of our applications.

We can now state our main theorem.

Theorem 2.4 (Main Theorem). *There exists a constant $C > 0$ such that the following is true. Let B be a finite set and let V be a linear subspace of functions $f : B \rightarrow \mathbb{Q}$. Assume that the following conditions hold for some integers $c_1, c_2, c_3 \geq 1$,*

1. *Divisibility: c_1 is the divisibility constant of V .*
2. *Boundedness of V : V has a c_2 -bounded integer basis in ℓ_∞ .*
3. *Boundedness of V^\perp : V^\perp has a c_3 -bounded integer basis in ℓ_1 .*
4. *Symmetry: for any $b_1, b_2 \in B$ there exists a symmetry π of V satisfying $\pi(b_1) = b_2$.*
5. *Constant functions: The constant functions belong to V .*

If

$$N \text{ is an integer multiple of } c_1 \text{ satisfying } \min(N, |B| - N) \geq C \cdot c_2 c_3^2 \dim(V)^6 \log(2c_3 \dim(V))^6 \quad (11)$$

then there exists a subset $T \subset B$ of size $|T| = N$ satisfying

$$\frac{1}{|T|} \sum_{t \in T} f(t) = \frac{1}{|B|} \sum_{b \in B} f(b) \quad \text{for all } f \text{ in } V. \quad (12)$$

A second goal of our work is to count the number of subsets T of a given size N which satisfy (12). To this end, we define a parameter $\rho(V)$ of the vector space V as follows. It is easiest to define $\rho(V)$ via a choice of basis for V but we stress that its value is independent of this choice. If $\phi : B \rightarrow \mathbb{Q}^A$ is such that the vectors (ϕ_a) , $a \in A$, form a basis for V , we define

$$\rho(V) := \frac{\det(\mathcal{L}(\phi))}{\sqrt{\det(\phi^t \phi)}}, \quad (13)$$

where in the numerator we mean the determinant of the lattice $\mathcal{L}(\phi)$ (i.e., the volume in the appropriate dimension of a fundamental parallelogram of the lattice generated by $\{\phi(b)\}$, $b \in B$) and in the denominator, the root of the determinant of the $A \times A$ matrix whose a, a' entry is $\sum_{b \in B} \phi(b)_a \phi(b)_{a'}$ (we denote by ϕ^t the transpose of ϕ). The definition takes a more symmetric form upon noting that the denominator is the determinant of the lattice in \mathbb{Q}^B generated by $\{\phi_a\}$, $a \in A$.

Theorem 2.5. *There exists a constant $C > 0$ such that the following is true. Let B be a finite set and let V be a linear subspace of functions $f : B \rightarrow \mathbb{Q}$. Assume that the conditions of Theorem 2.4 are satisfied with constants $c_1, c_2, c_3 \geq 1$ and that N satisfies (11). Then the number of subsets $T \subset B$ of size N which satisfy (12) equals*

$$\frac{\rho(V)}{(2\pi p(1-p))^{\frac{\dim(V)}{2}} p^N (1-p)^{|B|-N}} (1 + \delta)$$

where $p := \frac{N}{|B|}$ and $|\delta| \leq \frac{C \dim(V)^3 (\log(2c_2 \dim(V)))^{3/2}}{\sqrt{\min(N, |B|-N)}}$.

As explained before, the divisibility requirement in (11) is a necessary condition for the existence of a subset $T \subset B$ of size N satisfying (12). Our theorem says that when N is not too close to 0 or $|B|$ this condition is also sufficient, and gives a rather precise count of the number of such subsets.

Finally, we remark that our techniques yield a bit more. One can use them to show the existence, and estimate the number, of subsets T of a given size on which the average of functions $f \in V$ has a specified (small) displacement from the average over all of B . This extension is described in Sections 4.1 and 4.8.

3 Applications

In this section we apply our main theorem, Theorem 2.4, to prove the existence results for orthogonal arrays and t -designs, Theorems 1.1, 1.2, 1.3 and 1.6. It will be useful to introduce the following notation. For a map $\phi : B \rightarrow \mathbb{Q}^A$ and a vector $\gamma \in \mathbb{Q}^B$ we let

$$\phi(\gamma) = \sum_{b \in B} \gamma_b \phi(b) \in \mathbb{Q}^A.$$

We also define

$$\|\phi\|_\infty := \max_{b \in B, a \in A} |\phi(b)_a|.$$

We start by describing a certain condition which implies the boundedness condition for V^\perp and which will be useful in our applications to orthogonal arrays and t -designs.

3.1 Local decodability

In all of our applications it turns out that the most difficult condition to verify is that V^\perp has a bounded integer basis in ℓ_1 . This condition can be seen as an analog of the Low Density Parity Check (LDPC) notion coming from coding theory. We next introduce another condition which implies that V^\perp has a bounded integer basis in ℓ_1 , but which is sometimes easier to verify in practice. This condition is motivated by the notion of locally decodable codes in coding theory. Local decodability of codes is mainly studied in the context of codes defined over finite fields, see, e.g., [Yek11]. Here, we study an analog of these definitions for codes defined over the rationals.

Formally, we require that for some bounded integer basis (ϕ_a) , $a \in A$, of V , we may express a small multiple of the unit vectors (in the basis given by A) by short integer combinations of $\phi(b)$. The condition is also related to the notion of bi-orthogonal system in functional analysis.

Definition 3.1 (Local decodability). *A map $\phi : B \rightarrow \mathbb{Z}^A$ is locally decodable with bound c if there exists an integer $m \geq 1$ with $|m| \leq c$ and a set of vectors $(\gamma^a) \subset \mathbb{Z}^B$, $a \in A$, satisfying $\|\gamma^a\|_1 \leq c$ and*

$$\phi(\gamma^a) = m \cdot e^a \quad (a \in A),$$

where $e^a \in \{0, 1\}^A$ is the unit vector with 1 in coordinate a .

Claim 3.2. *If V has a basis of integer vectors $(\phi_a) \subset \mathbb{Z}^B$, $a \in A$, such that $\|\phi\|_\infty \leq c_2$ and ϕ is locally decodable with bound c_4 then V^\perp has a c_3 -bounded integer basis in ℓ_1 with $c_3 \leq 2c_2c_4|A|$.*

Proof. Let m and (γ^a) , $a \in A$, be as in Definition 3.1 for ϕ . Define the vectors $(\delta^b) \subset \mathbb{Z}^B$, $b \in B$ by

$$\delta^b := m \cdot u^b - \sum_{a \in A} \phi(b)_a \cdot \gamma^a,$$

where $u^b \in \{0, 1\}^B$ is the unit vector with 1 in coordinate b . We claim that the set $\{\delta^b : b \in B\}$ linearly spans V^\perp . First, note that $\delta^b \in V^\perp$ for all $b \in B$ since

$$\phi(\delta^b) = m \cdot \phi(u^b) - \sum_{a \in A} \phi(b)_a \cdot \phi(\gamma^a) = m \cdot \phi(b) - \sum_{a \in A} \phi(b)_a \cdot (m \cdot e_a) = 0.$$

We next argue that the rank of $\{\delta_b : b \in B\}$ is at least $|B| - |A|$, and hence they must span V^\perp . To see this, let $\Psi : \mathbb{Q}^B \rightarrow \mathbb{Q}^{|B|-|A|}$ be an arbitrary surjective linear map which sends $\{\gamma^a : a \in A\}$ to zero. Then $\{\delta_b : b \in B\}$ are mapped to a basis of $\mathbb{Q}^{|B|-|A|}$ by Ψ and hence their dimension is at least $|B| - |A|$ (in other words, the $\{\delta^b\}$ are a linear perturbation of the $B \times B$ identity matrix by a matrix whose rank is at most $|A|$). The bound on c_3 follows since

$$\|\delta_b\|_1 \leq m + \sum_{a \in A} |\phi(b)_a| \|\gamma^a\|_1 \leq c_4 + |A|c_2c_4 \leq 2c_2c_4|A| \quad (b \in B). \quad \square$$

3.2 Orthogonal arrays

We prove Theorems 1.1 and 1.2 in this subsection. We recall the relevant definitions from the introduction. A subset $T \subset [q]^n$ is an *orthogonal array of alphabet size q , length n and strength t* if it yields all strings of length t with equal frequency when restricted to any t coordinates. In other words, for any distinct indices $i_1, \dots, i_t \in [n]$ and any (not necessarily distinct) values $v_1, \dots, v_t \in [q]$,

$$|\{x \in T : x_{i_1} = v_1, \dots, x_{i_t} = v_t\}| = q^{-t}|T|.$$

Orthogonal arrays fit into our general framework as follows. We take $B := [q]^n$ and V to be the space spanned by all functions of the form

$$f_{(I,v)}(x_1, \dots, x_n) = \begin{cases} 1 & x_i = v_i \text{ for all } i \in I \\ 0 & \text{Otherwise} \end{cases},$$

with $I \subset [n]$ a subset of size t and $v \in [q]^I$. With this choice, a subset $T \subset B$ satisfying (1) is precisely an orthogonal array of alphabet size q , length n and strength t .

We shall now verify the conditions of Theorem 2.4 for V . We note that the sum of all the above $f_{(I,v)}$ is a constant function, thus verifying the constant functions condition. We continue with the symmetry condition. Fix $x \in [q]^n$ and consider the permutation $\pi_x \in S_B$ given by $\pi_x(b) = b + x \pmod{q}$, where we apply the modulo q coordinate-wise, and with the convention that it maps \mathbb{Z} to $[q]$. We will show that each π_x is a symmetry of V , which will establish the condition since the group $\{\pi_x : x \in [q]^n\}$ acts transitively on B . It suffices to show that for (I, v) of the above type, we have $f_{(I,v)} \circ \pi_x \in V$. Indeed,

$$(f_{(I,v)} \circ \pi_x)(b) = f_{(I,v)}(b + x \pmod{q}) = f_{(I,v')}(b) \in V,$$

where $v'_i = v_i - x_i \pmod{q}$ for $i \in I$.

To verify the remaining conditions in Theorem 2.4 we choose a convenient basis for V . The above set of functions $\{f_{(I,v)}\}$ is linearly dependent in general. Let

$$A := \{(I, v) : |I| \leq t, v \in [q-1]^I\},$$

and set $\phi_a = f_{(I,v)}$ for $a = (I, v) \in A$. Here, by $f_{(\emptyset, \emptyset)}$ we mean the constant one function. We will show that the (ϕ_a) , $a \in A$, form a basis for V , that the lattice $\mathcal{L}(\phi)$ which they generate equals \mathbb{Z}^A and that they are locally decodable. The remaining conditions of Theorem 2.4 will follow easily from these properties.

Claim 3.3. *The span of the functions $\{\phi_a\}_{a \in A}$ is V .*

Proof. Clearly $\phi_a \in V$ for all $a \in A$. To see that the $\{\phi_a\}_{a \in A}$ also span V , it suffices to show that any $f_{(I,v)}$ with $|I| \leq t$ and $v \in [q]^I$ is in the span of $\{\phi_a\}_{a \in A}$. We do this by induction on the number of elements in v which are equal to q . Let (I, v) have $|I| \leq t$ and $v \in [q]^I$. First, if $v \in [q-1]^I$ then $(I, v) \in A$ by definition. Now suppose $v \in [q]^I \setminus [q-1]^I$ and let $i_0 \in I$ be such that $v_{i_0} = q$. For $j \in [q-1]$ define v^j by $v^j_i = v_i$, $i \in I \setminus \{i_0\}$, and $v^j_{i_0} = j$. Define v' to be the restriction of v to $I \setminus \{i_0\}$. Then

$$f_{(I,v)} = f_{(I \setminus \{i_0\}, v')} - \sum_{j=1}^{q-1} f_{(I, v^j)}$$

and, by induction, the right hand side belongs to the linear span of $\{\phi_a\}_{a \in A}$. \square

Claim 3.4. *The map ϕ is locally decodable with bound 2^t and $m = 1$. Consequently, the (ϕ_a) , $a \in A$, form a basis for V and $\mathcal{L}(\phi) = \mathbb{Z}^A$.*

Proof. For $J \subset I \subset [n]$ and $v \in [q]^I$, we write $v|_J$ for the restriction of v to J . Define a partial order on A by letting $a' \leq a$, for $a = (I, v)$ and $a' = (I', v')$, if $I' \subset I$ and $v' = v|_{I'}$. For each $a = (I, v) \in A$, define an element $b^a \in [q]^n$ by

$$b_i^{(a)} := \begin{cases} v_i & \text{if } i \in I \\ q & \text{if } i \notin I \end{cases}.$$

The definition of ϕ implies that

$$\phi_{a'}(b^a) = 1_{(a' \leq a)} \quad (a, a' \in A).$$

Define for each $a = (I, v) \in A$ the vector $\gamma^a \in \mathbb{Z}^B$ by

$$\gamma_b^a := \begin{cases} (-1)^{|I|-|J|} & b = b^{(J, v|_J)} \text{ for some } J \subset I \\ 0 & \text{otherwise} \end{cases}.$$

We have $\|\gamma^a\|_1 = 2^{|I|} \leq 2^t$. We will show that $\phi(\gamma^a) = e^a$ for each $a \in A$, thereby establishing the local decodability claim. By the above, if $a = (I, v)$ and $a' = (I', v')$ then

$$\begin{aligned} \phi(\gamma^a)_{a'} &= \sum_{J \subset I} (-1)^{|I|-|J|} \phi_{a'}(b^{(J, v|_J)}) = \sum_{J \subset I} (-1)^{|I|-|J|} 1_{((I', v') \leq (J, v|_J))} = \\ &= 1_{(v' = v|_{I'})} \sum_{I' \subset J \subset I} (-1)^{|I|-|J|} = 1_{(v' = v|_{I'})} \sum_{j=0}^{|I|-|I'|} (-1)^j \binom{|I| - |I'|}{j} = \delta_{a, a'} \end{aligned}$$

as we wanted to show. Local decodability implies that the (ϕ_a) , $a \in A$, have full rank, and together with Claim 3.3 we deduce that they form a basis for V . Additionally, the fact that $\phi(\gamma^a) = e^a$ for every $a \in A$ implies that $\mathcal{L}(\phi) = \mathbb{Z}^A$. \square

Claim 3.5. *The divisibility constant of V equals q^t .*

Proof. It suffices to show that q^t is the smallest positive integer N for which

$$\frac{N}{|B|} \sum_{b \in B} \phi(b) \in \mathcal{L}(\phi)$$

Indeed, since $\mathcal{L}(\phi) = \mathbb{Z}^A$ by Claim 3.4 this follows by noting that

$$\frac{1}{|B|} \sum_{b \in B} \phi(b)_{(I, v)} = \frac{1}{q^n} |\{x \in [q]^n : x_i = v_i \ \forall i \in I\}| = q^{-|I|}. \quad \square$$

We are now in place to apply Theorem 2.4. The divisibility constant of V is $c_1 = q^t$. The $\{\phi_a\}$, $a \in A$, are a 1-bounded integer basis for V giving $c_2 = 1$. We have

$$\dim(V) = |A| = \sum_{i=0}^t \binom{n}{i} (q-1)^i \leq \binom{n}{t} q^t \leq \left(\frac{eqn}{t}\right)^t \quad (14)$$

with the next to last inequality following since V was defined as the span of $f_{(I, v)}$ with $|I| = t$ and $v \in [q]^I$. Since local decodability holds with bound $c_4 = 2^t$, we deduce from Claim 3.2 that V^\perp has a c_3 -bounded integer basis in ℓ_1 with $c_3 \leq 2c_2c_4|A| \leq 2\left(\frac{2eqn}{t}\right)^t$. Hence we establish the existence of an orthogonal array of alphabet size q , length n , strength t and size $|T| \leq \left(\frac{eqn}{t}\right)^{ct}$ for some universal constant $c > 0$.

Lastly, we aim to use Theorem 2.5 to count the number of orthogonal arrays of a given size. To this end we need only calculate $\rho(V)$. Observing that for our choice of ϕ we have $\det(\mathcal{L}(\phi)) = 1$ by Claim 3.4, we may apply formula (13) (with our choice of ϕ) to obtain

$$\rho(V) = \frac{\det(\mathcal{L}(\phi))}{\sqrt{\det(\phi^t \phi)}} = \frac{1}{\sqrt{\det(\phi^t \phi)}}.$$

Claim 3.6. $\det(\phi^t \phi) = q^{n \binom{n-1}{t} (q-1)^t}$.

It follows from the claim that

$$\rho(V) = q^{-n \binom{n-1}{t} (q-1)^t / 2}.$$

Together with the calculation of $\dim(V)$ in (14), Theorem 1.2 now follows from Theorem 2.5.

Proof of Claim 3.6. It will be useful to let n and t vary in the proof of the claim. Hence we shall write, for $n \geq 1$ and $0 \leq t \leq n$, $A(n, t)$ for A and $\phi(n, t)$ for ϕ . Denote $R(n, t) := \phi(n, t)^t \phi(n, t)$ (where $\phi(n, t)^t$ denotes the transpose of $\phi(n, t)$). Then $R(n, t)$ is an $A(n, t) \times A(n, t)$ matrix satisfying

$$R(n, t)_{(I, v), (I', v')} = |\{(x_1, \dots, x_n) \in [q]^n : x_i = v_i \forall i \in I \text{ and } x_{i'} = v_{i'} \forall i' \in I'\}|. \quad (15)$$

We also let

$$d_{n, t} := |A(n, t)| = \sum_{i=0}^t \binom{n}{i} (q-1)^i. \quad (16)$$

Define $a_{n, t} := \log_q(\det(R(n, t)))$ so that the claim states

$$a_{n, t} = n \binom{n-1}{t} (q-1)^t. \quad (17)$$

We first establish this fact when $t = 0$ or $n = t$. Indeed, if $t = 0$ we have $A(n, t) = \{(\emptyset, \emptyset)\}$ and $R(n, t)_{(\emptyset, \emptyset), (\emptyset, \emptyset)} = q^n$, proving that $a_{n, 0} = n$. If $n = t$, $|A(n, t)| = q^n$ and hence $\phi(n, t)$ is a *square* matrix with the property that the lattice spanned by its rows, by Claim 3.4, is $\mathbb{Z}^{A(n, t)}$. Thus $\det(\phi(n, t)) = 1$ and hence $\det(R(n, t)) = 1 = q^0$, verifying (17) in this case as well.

In the rest of the proof we will show that for any $n > t > 0$ we have

$$a_{n, t} = a_{n-1, t} + d_{n-1, t} + (q-1)a_{n-1, t-1} - d_{n-1, t-1}. \quad (18)$$

Noting that $d_{n-1, t} - d_{n-1, t-1} = \binom{n-1}{t} (q-1)^t$ by (16), the claim follows upon verifying that the $(a_{n, t})$ defined by (17) satisfy this recursion.

To prove (18), fix $n > t > 0$ and partition $A(n, t)$ as follows:

$$\begin{aligned} A(n, t)^0 &:= \{(I, v) \in A(n, t) : n \notin I\}, \\ A(n, t)^j &:= \{(I, v) \in A(n, t) : n \in I \text{ and } v_n = j\}, \quad j \in [q-1]. \end{aligned}$$

Observe that $A(n-1, t-1) \subset A(n-1, t) = A(n, t)^0$. Denote by $R(n, t)^{i, j}$ the sub-matrix of $R(n, t)$ whose rows are indexed by $A(n, t)^i$ and whose columns are indexed by $A(n, t)^j$. By (15) we have

$$R(n, t)^{i, j} = \begin{cases} qR(n-1, t) & i = j = 0 \\ 0 & i, j \in [q-1], i \neq j \\ R(n-1, t-1) & i, j \in [q-1], i = j \end{cases} \quad (19)$$

where in the second case we mean the 0 matrix, and in the third case we identified $a = (I, v) \in A(n, t)^j$ with $a' = (I', v') \in A(n-1, t-1)$ defined by letting $I' = I \setminus \{n\}$ and $v'_i = v_i$ for $i \in I'$. Summarizing these equalities we have

$$R(n, t) = \begin{pmatrix} qR(n-1, t) & R(n, t)^{0,1} & R(n, t)^{0,2} & \dots & R(n, t)^{0, q-1} \\ R(n, t)^{1,0} & R' & 0 & \dots & 0 \\ R(n, t)^{2,0} & 0 & R' & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(n, t)^{q-1,0} & 0 & 0 & \dots & R' \end{pmatrix},$$

where the i, j cell in the displayed matrix corresponds to the sub-matrix indexed by $A(n, t)^i$ and $A(n, t)^j$, and where we abbreviated $R' := R(n-1, t-1)$ for display purposes. In addition, (15) implies that if $a_1 \in A(n, t)^{j_1}$ and $a_2 \in A(n, t)^{j_2}$ then

$$R(n, t)_{a_1, a_2} = \begin{cases} \frac{1}{q}R(n, t)_{a_1, a'_2} & j_1 = 0, j_2 \neq 0 \\ R(n, t)_{a_1, a'_2} & j_1, j_2 \neq 0, j_1 = j_2 \end{cases}. \quad (20)$$

Thus, if we subtract from each of the columns indexed by $a_2 \in A(n, t) \setminus A(n, t)^0$ the corresponding column indexed by $a'_2 \in A(n, t)^0$ times $\frac{1}{q}$ we obtain the matrix $\tilde{R}(n, t)$ satisfying

$$\tilde{R}(n, t) = \begin{pmatrix} qR(n-1, t) & 0 & 0 & \cdots & 0 \\ R(n, t)^{1,0} & (1 - \frac{1}{q})R' & -\frac{1}{q}R' & \cdots & -\frac{1}{q}R' \\ R(n, t)^{2,0} & -\frac{1}{q}R' & (1 - \frac{1}{q})R' & \cdots & -\frac{1}{q}R' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(n, t)^{q-1,0} & -\frac{1}{q}R' & -\frac{1}{q}R' & \cdots & (1 - \frac{1}{q})R' \end{pmatrix},$$

Denoting by I the identity matrix and by $\mathbf{1}$ the square matrix containing all ones, both of dimension $q-1$, it follows that

$$\begin{aligned} \det(R(n, t)) &= \det(\tilde{R}(n, t)) = \det(qR(n-1, t)) \det\left(\left(I - \frac{1}{q}\mathbf{1}\right) \otimes R(n-1, t-1)\right) = \\ &= q^{a_{n-1,t} + d_{n-1,t} + (q-1)a_{n-1,t-1}} \det\left(I - \frac{1}{q}\mathbf{1}\right)^{d_{n-1,t-1}} = q^{a_{n-1,t} + d_{n-1,t} + (q-1)a_{n-1,t-1} - d_{n-1,t-1}}, \end{aligned}$$

where we used that if A is an $m \times m$ matrix and B is a $k \times k$ matrix then $\det(A \otimes B) = \det(A)^k \det(B)^m$, and where we calculated $\det(I - \frac{1}{q}\mathbf{1}) = \frac{1}{q}$ since the only non-zero eigenvalue of $\mathbf{1}$ is $q-1$. This establishes the recursion (18) and finishes the proof of the claim. \square

3.3 Designs

We prove Theorems 1.3 and 1.4 in this subsection. We recall the relevant definitions from the introduction. A (simple) t -(v, k, λ) *design*, or t -design for short, is a family of distinct subsets of $[v]$, where each set is of size k , such that each t elements belong to exactly λ sets. In other words, denoting by $\begin{bmatrix} v \\ k \end{bmatrix}$ the family of all subsets of $[v]$ of size k , a set $T \subset \begin{bmatrix} v \\ k \end{bmatrix}$ is a t -design if for any distinct elements $i_1, \dots, i_t \in [v]$,

$$|\{s \in T : i_1, \dots, i_t \in s\}| = \frac{\binom{k}{t}}{\binom{v}{t}} |T| = \lambda. \quad (21)$$

Our general framework includes t -designs as follows. We take B to be $\begin{bmatrix} v \\ k \end{bmatrix}$ and V to be the space spanned by all functions of the form

$$f_a(b) = \begin{cases} 1 & a \subset b \\ 0 & \text{Otherwise} \end{cases},$$

with $a \in \begin{bmatrix} v \\ t \end{bmatrix}$. With this choice, a subset $T \subset B$ satisfying (1) is precisely a simple t -(v, k, λ) design. We choose $A = \begin{bmatrix} v \\ t \end{bmatrix}$, set $\phi_a = f_a$ and define a map $\phi : B \rightarrow \mathbb{Z}^A$ by $\phi(b)_a = \phi_a(b)$.

Fix $v \geq 1$ and $1 \leq t \leq k \leq v$. We assume without loss of generality that $k \leq v-t$ since if $k > v-t$ we have $|B| = \binom{v}{k} \leq \binom{v}{t}$ and hence Theorem 1.3 holds trivially and Theorem 1.4

holds vacuously (by taking the complete design in Theorem 1.3, and by noting that necessarily $\binom{v}{k} - N \leq \binom{v}{t}$ in Theorem 1.4). This assumption will be needed shortly to show that $\{\phi_a\}$, $a \in A$, form a basis for V (that is, that they are linearly independent).

We shall now verify the conditions of Theorem 2.4 for V . First, the boundedness condition for V trivially holds with constant $c_2 = 1$ by our choice of ϕ . Second, we observe that $\sum_{a \in A} \phi_a$ is the vector with all coordinates equal to $\binom{k}{t}$. Hence the constant functions assumption is satisfied. Third, to establish the symmetry condition let $\pi \in S_{[v]}$ be a permutation on $[v]$. π acts in a natural way on B (by permuting k -sets) and on A (by permuting t -sets). We have that

$$(\phi_a \circ \pi)(b) = \phi_a(\pi(b)) = 1_{a \subset \pi(b)} = 1_{\pi^{-1}(a) \subset b} = \phi_{\pi^{-1}(a)}(b),$$

and in particular $\phi_a \circ \pi \in V$ for all $a \in A$. The action of $S_{[v]}$ on B is transitive, from which the symmetry condition follows. We continue to find the divisibility constant of V . We need the following result of Wilson [Wil73] and Graver and Jurkat [GJ73].

Theorem 3.7. *The vector in \mathbb{Q}^A all of whose coordinates equal λ belongs to $\mathcal{L}(\phi)$ if and only if*

$$\binom{v-s}{t-s} \lambda \equiv 0 \pmod{\binom{k-s}{t-s}} \quad \text{for all } 0 \leq s \leq t.$$

Define $c_1 \geq 1$ to be the minimal integer such that

$$\binom{k}{s} c_1 \equiv 0 \pmod{\binom{v}{s}} \quad \text{for all } 1 \leq s \leq t. \quad (22)$$

We claim that c_1 is the divisibility constant of V . Indeed, since all the coordinates of $\frac{N}{|B|} \sum_{b \in B} \phi(b)$ equal $N \binom{v-t}{k-t} / \binom{v}{k} = N \binom{k}{t} / \binom{v}{t}$ we see that this vector belongs to $\mathcal{L}(\phi)$ if and only if

$$\frac{\binom{v-s}{t-s} \binom{k}{t}}{\binom{k-s}{t-s} \binom{v}{t}} N = \frac{\binom{k}{s}}{\binom{v}{s}} N \in \mathbb{Z} \quad \text{for all } 0 \leq s \leq t.$$

The case $s = 0$ simply means that $N \in \mathbb{Z}$, thus a comparison with (22) verifies that c_1 is the divisibility constant of V .

It is useful to have a simple upper bound for c_1 . Define

$$\text{lcm}(t) := \text{lcm} \left\{ \binom{t}{s} : 0 \leq s \leq t \right\}. \quad (23)$$

Observing that

$$\frac{\binom{k}{s}}{\binom{v}{s}} \cdot \binom{v}{t} \binom{t}{s} = \binom{v-s}{v-t} \binom{k}{s} \in \mathbb{Z} \quad \text{for all } 1 \leq s \leq t$$

we deduce that $c_1 \leq \binom{v}{t} \text{lcm}(t)$. We note that [Far09] shows that $\log(\text{lcm}(t))$ is asymptotic to t and mentions effective bounds for it. The next claim provides a simple self-contained proof of a weaker bound which suffices for our needs. It follows from the claim that $c_1 \leq \binom{v}{t} 4^t \leq (4ev/t)^t$.

Claim 3.8. $\text{lcm}(t) \leq 4^t$ for $t \geq 1$.

Proof. Assume by induction that the claim holds up to t (checking also that it holds for $t = 1$) and let us prove it for t . For each $0 \leq s \leq \lfloor \frac{t}{2} \rfloor$ we have

$$\binom{t}{s} = \frac{t!}{s!(t-s)!} = \frac{t!}{\lceil \frac{t}{2} \rceil! s! (t-s)!} = \frac{t!}{\lceil \frac{t}{2} \rceil! s!} \prod_{i=\lceil \frac{t}{2} \rceil+1}^{t-s} \frac{1}{i}.$$

Since the product of every m consecutive integers is divisible by $m!$ (since $\binom{a}{m}$ is an integer for every a), using the symmetry of the binomial coefficients and the induction hypothesis we have

$$\begin{aligned} \text{lcm} \left(\left\{ \binom{t}{s} : 0 \leq s \leq t \right\} \right) &= \text{lcm} \left(\left\{ \frac{t!}{\lceil \frac{t}{2} \rceil! s!} \prod_{i=\lceil \frac{t}{2} \rceil+1}^{t-s} \frac{1}{i} : 0 \leq s \leq \left\lfloor \frac{t}{2} \right\rfloor \right\} \right) \leq \\ &\leq \text{lcm} \left(\left\{ \frac{t!}{\lceil \frac{t}{2} \rceil! s! (\lfloor \frac{t}{2} \rfloor - s)!} : 0 \leq s \leq \left\lfloor \frac{t}{2} \right\rfloor \right\} \right) = \\ &= \binom{t}{\lfloor \frac{t}{2} \rfloor} \text{lcm} \left(\left\{ \binom{\lfloor \frac{t}{2} \rfloor}{s} : 0 \leq s \leq \left\lfloor \frac{t}{2} \right\rfloor \right\} \right) \leq 2^t 4^{\lfloor \frac{t}{2} \rfloor} \leq 4^t. \quad \square \end{aligned}$$

Thus, to verify the conditions of Theorem 2.4, it remains only to verify the boundedness assumption for V^\perp . We do so by applying the local decodability claim, Claim 3.2, to the map ϕ . Let $a \in A = \begin{bmatrix} v \\ t \end{bmatrix}$. Let $u \in \begin{bmatrix} v \\ k+t \end{bmatrix}$ be a any set of size $k+t$ such that $a \subset u$ (here we use our assumption that $k \leq v-t$). We denote by $\begin{bmatrix} u \\ k \end{bmatrix} \subset B$ the family of subsets of u of size k . Define $\gamma_{a,u} \in \mathbb{Z}^B$ as

$$\gamma_{a,u} := \sum_{s=0}^t \sum_{b \in \begin{bmatrix} u \\ k \end{bmatrix} : |a \cap b| = s} (-1)^{t-s} \frac{s!(k-s-1)!}{(k-t-1)!} \cdot u_b,$$

where $u_b \in \{0,1\}^B$ is the unit vector with 1 in coordinate b .

Claim 3.9. $\phi(\gamma_{a,u}) = \frac{k!}{(k-t)!} \cdot e_a$ for all $a \in A$.

Note that the claim implies, in particular, that the (ϕ_a) , $a \in A$, are independent and thus

$$\dim(V) = |A| = \binom{v}{t}. \quad (24)$$

The next technical claim is used in the proof of Claim 3.9. We set $\binom{n}{m} = 0$ whenever $n < m$.

Claim 3.10. Let $a > b \geq 0$ and $c \geq 0$. Then

$$\sum_{i=0}^a (-1)^i \binom{a}{i} \binom{c+i}{b} = 0.$$

Proof. Let $f(a,b,c) = \sum_{i=0}^a (-1)^i \binom{a}{i} \binom{c+i}{b}$. If $b, c > 0$ we have $\binom{c+i}{b} = \binom{c-1+i}{b} + \binom{c-1+i}{b-1}$ and hence $f(a,b,c) = f(a,b,c-1) + f(a,b-1,c-1)$. So, it is enough to verify the claim whenever $b = 0$ or $c = 0$. If $b = 0$ then $f(a,0,c) = \sum_{i=0}^a (-1)^i \binom{a}{i} = 0$ since $a \geq 1$. If $c = 0$ then $f(a,b,0) = \sum_{i=b}^a (-1)^i \binom{a}{i} \binom{i}{b} = \binom{a}{b} \sum_{i=b}^a (-1)^i \binom{a-b}{i-b} = 0$. \square

Proof of Claim 3.9. It is clear from the definition that $\phi(\gamma_{a,u})_{a'} = 0$ if $a' \not\subset u$. So, we restrict our attention to $a' \subset u$. For $a' = a$ the contribution is only from sets with $s = |a \cap b| = t$, of which there are $\binom{k}{k-t}$, and hence

$$\phi(\gamma_{a,u})_a = \binom{k}{k-t} t! = \frac{k!}{(k-t)!}.$$

We now need to verify that $\phi(\gamma_{a,u})_{a'} = 0$ for all $a' \subset u$, $a' \neq a$. Let us denote $\ell = |a \cap a'|$ where $0 \leq \ell < t$. The contribution to $\phi(\gamma_{a,u})_{a'}$ comes only from sets b for which $a' \subset b$. The number of

these sets with $|a \cap b| = s$ is $\binom{t-\ell}{s-\ell} \binom{k-t+\ell}{k-t-s+\ell} = \binom{t-\ell}{t-s} \binom{k-t+\ell}{s}$. Note, moreover, that $s \geq \ell$. We have

$$\begin{aligned} \phi(\gamma_{a,u})_{a'} &= \sum_{s=\ell}^t \binom{t-\ell}{t-s} \binom{k-t+\ell}{s} \cdot (-1)^{t-s} \frac{s!(k-s-1)!}{(k-t-1)!} \\ &= \frac{(t-\ell-1)!(k-t+\ell)!}{(k-t-1)!} \sum_{s=\ell}^t (-1)^{t-s} \binom{t-\ell}{t-s} \binom{k-s-1}{t-\ell-1} \\ &= \frac{(t-\ell-1)!(k-t+\ell)!}{(k-t-1)!} \sum_{i=0}^{t-\ell} (-1)^i \binom{t-\ell}{i} \binom{k-t-1+i}{t-\ell-1}. \end{aligned}$$

Recalling that $t-\ell > 0$ we now apply Claim 3.10 with $a = t-\ell, b = t-\ell-1, c = k-t-1$ and conclude that $\phi(\gamma_{a,u})_{a'} = 0$. \square

In order to obtain tight bounds, we will divide $\gamma_{a,u}$ by a factor common to all the coefficients appearing in it. Note that

$$\frac{s!(k-s-1)!}{(k-t-1)!} = \binom{k-s-1}{k-t-1} \binom{t}{s}^{-1} t!.$$

and hence

$$\gamma'_{a,u} := \frac{\text{lcm}(t)}{t!} \cdot \gamma_{a,u} \in \mathbb{Z}^B \quad (25)$$

We continue to show that ϕ is locally decodable. We have

$$\phi(\gamma'_{a,u}) = \binom{k}{t} \text{lcm}(t) \cdot e_a$$

and we recall that $\binom{k}{t} \text{lcm}(t) \leq \binom{k}{t} 4^t$ by Claim 3.8. To bound $\|\gamma'_{a,u}\|_1$ observe that the number of $b \in \binom{[u]}{k}$ for which $|a \cap b| = s$ is $\binom{t}{s} \binom{k}{k-s} = \binom{t}{s} \binom{k}{s}$; and $\frac{s!(k-s-1)!}{t!(k-t-1)!} = \binom{k-1}{t} / \binom{k-1}{s}$. Hence

$$\|\gamma'_{a,u}\|_1 = \text{lcm}(t) \sum_{s=0}^t \binom{t}{s} \binom{k}{s} \frac{\binom{k-1}{t}}{\binom{k-1}{s}} \leq 4^t \frac{k}{k-t} \binom{k-1}{t} \sum_{s=0}^t \binom{t}{s} = 8^t \binom{k}{t}$$

implying that ϕ is locally decodable with bound $c_4 = 8^t \binom{k}{t} \leq (8e \cdot k/t)^t$. Finally, since $|A| = \binom{v}{t} \leq (ev/t)^t$ we obtain from Claim 3.2 that V^\perp has a c_3 -bounded integer basis in ℓ_1 with $c_3 \leq 2c_2 c_4 |A| \leq (4ev/t)^{2t}$.

We have verified the conditions of Theorem 2.4 with $|A| \leq (ev/t)^t, c_1 \leq (4ev/t)^t, c_2 = 1, c_3 \leq (4ev/t)^{2t}$ and thus we establish the existence of a simple t -(v, k, λ) design of size $|T| \leq (cv/t)^{ct}$ for some universal constant $c > 0$, proving Theorem 1.3.

We turn to estimate the number of designs of a given size via Theorem 2.5. To this end we consider ϕ as a $B \times A$ matrix and need to calculate the parameter $\rho(V)$ from (13). We rely on a theorem of Wilson [Wil90] giving a diagonal form of ϕ^t and on a theorem of Bapat [Bap00] calculating the eigenvalues of $\phi^t \phi$.

Theorem 3.11. [Wil90, Theorem 2] *There exist a $\binom{v}{t} \times \binom{v}{t}$ matrix E and a $\binom{v}{k} \times \binom{v}{k}$ matrix F , both with integer entries and satisfying $|\det(E)| = |\det(F)| = 1$, such that $M := E\phi^t F$ has $M_{ij} = 0$ if $i \neq j$ and has diagonal entries $\binom{k-s}{t-s}$ with multiplicity $\binom{v}{s} - \binom{v}{s-1}$ for $0 \leq s \leq t$ (with $\binom{v}{-1} := 0$).*

Theorem 3.12. [Bap00, Theorem 4] *The eigenvalues of $\phi^t \phi$ are $\binom{k-s}{t-s} \binom{v-t-s}{k-t}$ with multiplicity $\binom{v}{s} - \binom{v}{s-1}$ for $0 \leq s \leq t$ (with $\binom{v}{-1} := 0$).*

The two theorems immediately imply that

$$\rho(V) = \frac{\det(\mathcal{L}(\phi))}{\sqrt{\det(\phi^t \phi)}} = \frac{\prod_{s=0}^t \binom{k-s}{t-s} \binom{v}{s} - \binom{v}{s-1}}{\prod_{s=0}^t \left[\binom{k-s}{t-s} \binom{v-t-s}{k-t} \right]^{\frac{1}{2} \left(\binom{v}{s} - \binom{v}{s-1} \right)}} = \prod_{s=0}^t \left[\frac{\binom{k-s}{t-s}}{\binom{v-t-s}{k-t}} \right]^{\frac{1}{2} \left(\binom{v}{s} - \binom{v}{s-1} \right)}.$$

Theorem 1.4 now follows by an application of Theorem 2.5 (recalling that $\dim(V) = \binom{v}{t}$ by (24)).

We remark briefly on a possible approach to proving Theorems 3.11 and 3.12 via representation theory (see Section 3.4.1 for some background), though we neither require nor develop this approach here. One may naturally identify the set $\left[\begin{smallmatrix} v \\ k \end{smallmatrix} \right]$ with tabloids of shape $(v-k, k)$ as the numbers appearing in the second row of the tabloid. Thus, \mathbb{R}^B may be identified with the Young module $U_{(v-k, k)}$. Similarly, \mathbb{R}^A may be identified with $U_{(v-t, t)}$. For any m , the decomposition of $U_{(v-m, m)}$ into irreducible representations is $U_{(v-m, m)} = \oplus_{s=0}^m V_{(v-s, s)}$. Since ϕ intertwines the action of S_n on $U_{(v-k, k)}$ and $U_{(v-t, t)}$, Schur's lemma implies that ϕ and ϕ^t are diagonal in the basis of irreducible representations for $U_{(v-k, k)}$ and $U_{(v-t, t)}$, and act as scalars from $V_{(v-s, s)}$ to itself, $0 \leq s \leq t$. Finally, for each s , the scalars appearing in these actions can be determined by considering the action on some particular vector in $V_{(v-s, s)}$. Choosing bases appropriately one obtains that the scalar for the ϕ action is $\binom{v-t-s}{k-t}$ and the scalar for the ϕ^t action is $\binom{k-s}{t-s}$, both with multiplicity $\dim(V_{(v-s, s)}) = \binom{v}{s} - \binom{v}{s-1}$.

3.4 t -wise permutations

We prove Theorem 1.6 in this subsection. We recall the relevant definitions from the introduction. A family of permutations $T \subset S_n$ is called a t -wise permutation if its action on any t -tuple of elements is uniform. In other words, for any distinct elements $i_1, \dots, i_t \in [n]$ and distinct elements $j_1, \dots, j_t \in [n]$,

$$|\{\pi \in T : \pi(i_1) = j_1, \dots, \pi(i_t) = j_t\}| = \frac{1}{n(n-1) \cdots (n-t+1)} |T|.$$

Our general framework includes t -wise permutations as follows. We first set notations. Let $[n]_t := \{(i_1, \dots, i_t) : i_1, \dots, i_t \in [n] \text{ distinct}\}$ denote the family of t -tuples of distinct elements. For $I = (i_1, \dots, i_t) \in [n]_t$ and $\pi \in S_n$ define $\pi(I) := (\pi(i_1), \dots, \pi(i_t)) \in [n]_t$. We take $B = S_n$ and W to be the space spanned by all functions of the form

$$f_{I, J}(\pi) = 1_{\pi(I)=J}. \quad (26)$$

where $I, J \in [n]_t$ (we changed notation for the subspace from V to W in this section to avoid confusion with notations arising later which are related to the representation theory of the symmetric group). With this choice, a subset $T \subset B$ satisfying (1) is precisely a t -wise permutation. We now establish the conditions of Theorem 2.4. We will show that:

1. The divisibility constant of W is $c_1 = \frac{n!}{(n-t)!}$.
2. W has a c_2 -bounded integer basis in ℓ_∞ with $c_2 = 1$.
3. W^\perp has a c_3 -bounded integer basis in ℓ_1 with $c_3 = (t+2)!$.

4. The group S_n acts on W transitively.
5. The space W contains the constant functions.
6. The dimension of W equals the number of permutations in S_n with longest increasing subsequence of length at least $n - t$. It satisfies $\dim(W) \leq |[n]_t|^2 \leq n^{2t}$.

With these conditions, Theorem 2.4 immediately implies Theorem 1.6.

A few conditions are easy to verify. First, W contains the constant functions since for each $I \in [n]_t$, $\sum_{J \in [n]_t} f_{I,J}$ is the constant function 1. Second, any spanning subset of the functions $f_{I,J}$ is an integer basis for W with ℓ_∞ norm $c_2 = 1$. Third, observe that S_n acts naturally on $B = S_n$ by composition. Each $\sigma \in S_n$ is a symmetry of W since for any $I, J \in [n]_t$,

$$(f_{I,J} \circ \sigma)(\pi) = f_{I,J}(\sigma(\pi)) = 1_{\sigma(\pi(I))=J} = 1_{\pi(I)=\sigma^{-1}(J)} = f_{I,\sigma^{-1}(J)}(\pi) \in W.$$

The action of S_n on B is transitive, from which the symmetry condition follows. Fourth, $\dim(W) \leq |[n]_t|^2$ since the $(f_{I,J})$ span W .

In order to find the divisibility constant of W and establish that W^\perp is spanned by integer vectors with small ℓ_1 norm, we will need some basic facts regarding the irreducible representations of the symmetric group, which we describe next. We will follow the notation of [FH91, Chapter 4], but also refer the reader to [Jam78] for the necessary background.

3.4.1 Irreducible representations of the symmetric group

Partitions. A *partition* λ of n is a vector $(\lambda_1, \dots, \lambda_\ell)$ of some length ℓ such that $\lambda_1 \geq \dots \geq \lambda_\ell \geq 1$ and $\sum_{i=1}^\ell \lambda_i = n$. If referring to λ_i for $i > \ell$ we adapt the convention that $\lambda_i = 0$ for such i . The *conjugate partition* λ' is defined as $(\lambda'_1, \dots, \lambda'_m)$ where $m = \lambda_1$ and $\lambda'_i := |\{j : \lambda_j \geq i\}|$. The *dominance partial order* \supseteq on partitions is defined by

$$\lambda \supseteq \mu \quad \text{if and only if} \quad \sum_{i=1}^j \lambda_i \geq \sum_{i=1}^j \mu_i \quad \text{for all } j \geq 1.$$

We let \geq stand for the *lexicographic total order* on partitions. It is well-known that the lexicographic order extends the dominance order, and that $\lambda \supseteq \mu$ if and only if $\mu' \supseteq \lambda'$. We denote by \mathcal{P}_n the set of all partitions of n .

Young diagrams, tableaux and tabloids. Let $\lambda \in \mathcal{P}_n$. Associated with it is the *Young diagram* of shape λ (in English notation). A *tableau* of shape λ is a filling of the Young diagram of shape λ with the integers 1 to n , with each integer occurring once. We say that two tableaux are *row-equivalent* if they are the same up to the order of the numbers in each row. A *tabloid* of shape λ is an equivalence class of tableaux for the row-equivalence relation. We denote the tabloid associated to the tableau T by $[T]$. S_n acts on tableaux with the permutation action on the numbers in each tableau. This induces a corresponding action on tabloids. The *column stabilizer* of a tableau T of shape λ is the subgroup $Q_T \leq S_n$ of permutations preserving the columns of T .

Irreducible representations, Young modules and Kostka numbers. The irreducible representations (over \mathbb{C} or \mathbb{Q}) of the symmetric group S_n are in correspondence with shapes $\lambda \in \mathcal{P}_n$. We denote by V_λ the irreducible representation corresponding to λ . We let U_μ , sometimes called the Young module or permutation module of shape μ , be the module whose basis is all tabloids of shape μ , equipped with the action of S_n on tabloids. Each Young module U_μ is isomorphic to a sum of irreducible representations. The *Kostka number* $K_{\lambda,\mu}$ is the multiplicity of the irreducible representation V_λ in U_μ . It is known that $K_{\lambda,\mu} = 0$ unless $\lambda \supseteq \mu$ and $K_{\lambda,\lambda} = 1$.

The group algebra and Fourier analysis. We denote by $\mathbb{C}S_n$ the group algebra of S_n , the set of functions $f : S_n \rightarrow \mathbb{C}$ endowed with the product

$$(f * g)(\pi) = \sum_{\sigma \in S_n} f(\sigma)g(\sigma^{-1}\pi).$$

We fix once and for all a matrix representation for each irreducible representation V_λ . Then the functions $V_\lambda(\cdot)_{i,j} : S_n \rightarrow \mathbb{C}$ for $\lambda \in \mathcal{P}_n$ and $1 \leq i, j \leq \dim(V_\lambda)$ are linearly independent and span $\mathbb{C}S_n$. Moreover, extending $V_\lambda(\cdot)$ linearly to all of $\mathbb{C}S_n$,

$$V_\lambda(f * g) = V_\lambda(f)V_\lambda(g) \quad \text{for } f, g \in \mathbb{C}S_n \text{ and } \lambda \in \mathcal{P}_n.$$

In addition, $f = 0$ if and only if $V_\lambda(f) = 0$ for all $\lambda \in \mathcal{P}_n$.

3.4.2 Two bases for W and the divisibility constant

Define W' to be the span of the functions $\{V_\lambda(\cdot)_{i,j} : \lambda_1 \geq n - t, 1 \leq i, j \leq \dim(V_\lambda)\}$. As all the $(V_\lambda(\cdot)_{i,j})$ are linearly independent, we have

$$\dim(W') = \sum_{\lambda \in \mathcal{P}_n : \lambda_1 \geq n-t} \dim(V_\lambda)^2. \quad (27)$$

In this section we show that $W = W'$ and define a combinatorial basis which is useful in determining the divisibility constant of W .

Claim 3.13. $W \subseteq W'$.

Proof. It suffices to show that $f_{I,J} \in W'$ for all $I, J \in [n]_t$. Consider the representation R_t of the action of S_n on t -tuples, defined by

$$R_t(\pi)_{I,J} = 1_{\pi(I)=J} = f_{I,J}(\pi) \quad \text{for } \pi \in S_n \text{ and } I, J \in [n]_t.$$

It is simple to see that R_t is isomorphic to the Young module U_μ for $\mu = (n - t, 1, 1, \dots, 1) \in \mathcal{P}_n$. Indeed, this follows by identifying each $I \in [n]_t$ with the tabloid of shape μ having the elements of I , in order, as the elements in its first column at rows $2, \dots, t + 1$. By definition of the Kostka number $K_{\lambda,\mu}$, if the irreducible representation V_λ appears in the decomposition of U_μ then $\lambda \succeq \mu$, which occurs if and only if $\lambda_1 \geq n - t$. In conclusion, the decomposition of R_t into irreducible representations contains only V_λ with $\lambda_1 \geq n - t$. Passing to the basis of irreducible representations, we conclude that $f_{I,J} \in W'$ for each $I, J \in [n]_t$, as required. \square

It is not difficult to use the same representation-theoretic methods to show that $W = W'$, but we proceed by a different route in order to identify also the divisibility constant of W .

For $\sigma \in S_n$, let $\text{LIS}(\sigma)$ denote the length of the longest increasing subsequence of σ . Define

$$A := \{\sigma \in S_n : \text{LIS}(\sigma) \geq n - t\}. \quad (28)$$

Claim 3.14. $|A| = \dim(W')$.

Proof. The Robinson-Schensted correspondence [Rob38, Sch61] shows that

$$|\{\sigma \in S_n : \text{LIS}(\sigma) = r\}| = \sum_{\lambda \in \mathcal{P}_n : \lambda_1 = r} \dim(V_\lambda)^2.$$

Summing over $n - t \leq r \leq n$ and comparing with (27) concludes the proof. \square

We now define a set of functions (f_σ) , $\sigma \in A$, forming a basis of W . For each $\sigma \in A$, let $S(\sigma) \subseteq [n]$ be the indices of an (arbitrary) increasing subsequence in σ of length $\text{LIS}(\sigma)$. That is, $S(\sigma) = (i_1, \dots, i_{\text{LIS}(\sigma)})$ for some indices satisfying $i_{j+1} > i_j$ and $\sigma(i_{j+1}) > \sigma(i_j)$. Define

$$f_\sigma(\pi) = \begin{cases} 1 & \pi(j) = \sigma(j) \quad \forall j \notin S(\sigma) \\ 0 & \text{otherwise} \end{cases}. \quad (29)$$

It is clear that the functions f_σ are in W since $|S(\sigma)| \geq n - t$. Let \succeq stand for the lexicographic order on S_n . I.e., $\pi \succ \sigma$ if there exists a j such that $\pi(i) = \sigma(i)$ for all $i < j$ and $\pi(j) > \sigma(j)$.

Lemma 3.15. *For each $\sigma \in A$, $f_\sigma(\sigma) = 1$ and if $f_\sigma(\pi) = 1$ then $\pi \succeq \sigma$.*

Proof. $f_\sigma(\sigma) = 1$ by the definition of f_σ . Suppose that $f_\sigma(\pi) = 1$. Then $\pi(i) = \sigma(i)$ for every $i \notin S(\sigma)$ and $\{\pi(i) : i \in S(\sigma)\} = \{\sigma(i) : i \in S(\sigma)\}$. Since $S(\sigma)$ are the indices of an increasing subsequence in σ it follows that π appears after σ in the lexicographic order. \square

Now let $\phi : B \rightarrow \mathbb{Z}^A$ be the matrix whose columns are the f_σ . It follows from Lemma 3.15 that when the rows and columns are sorted by the lexicographic order on permutations, then ϕ is in column-echelon form. Consequently, the columns of ϕ are linearly independent and hence $\dim(W) \geq |A|$. Combining this fact with Claims 3.13 and 3.14 shows that $W = W'$ and that the columns of ϕ form a basis for W . In addition, the column-echelon form of ϕ implies that $\mathcal{L}(\phi) = \mathbb{Z}^A$. Now, if $\sigma \in A$ has $\text{LIS}(\sigma) = n - \ell$ for some $0 \leq \ell \leq t$ then

$$\frac{1}{|B|} \sum_{\pi \in B} \phi(\pi)_\sigma = \frac{1}{n!} |\{\pi \in S_n : \pi(j) = \sigma(j) \quad \forall j \notin S(\sigma)\}| = \frac{(n - \ell)!}{n!}.$$

Hence $\frac{N}{|B|} \sum_{b \in B} \phi(b) \in \mathbb{Z}^A = \mathcal{L}(\phi)$ if and only if N is a multiple of $\frac{n!}{(n-t)!}$, implying that $c_1 = \frac{n!}{(n-t)!}$ is the divisibility constant of W .

We end the section by giving an alternate characterization of W .

Claim 3.16. *A function $f \in \mathbb{C}S_n$ satisfies $f \in W$ if and only if $V_\lambda(f) = 0$ for every $\lambda \in \mathcal{P}_n$ with $\lambda_1 \leq n - t - 1$.*

Proof. We first recall the orthogonality relations for irreducible representations which state that

$$\sum_{\pi \in S_n} V_\mu(\pi^{-1})_{i,j} V_\lambda(\pi)_{k,\ell} = \frac{n!}{\dim(V_\lambda)} \delta_{\lambda,\mu} \delta_{i,k} \delta_{j,\ell} \quad \lambda, \mu \in \mathcal{P}_n, 1 \leq i, j \leq \dim(V_\lambda), 1 \leq k, \ell \leq \dim(V_\mu). \quad (30)$$

We continue by observing that $f \in W$ if and only if $\tilde{f} \in W$ where \tilde{f} is defined by $\tilde{f}(\pi) = f(\pi^{-1})$. Indeed, $f \in W$ if and only if $f = \sum_{I,J \in [n]_t} \alpha_{I,J} f_{I,J}$ for the functions $f_{I,J}$ defined by (26) and some coefficients $\alpha_{I,J}$. The observation now follows since $f_{I,J}(\pi^{-1}) = f_{J,I}(\pi)$.

Now let $f \in \mathbb{C}S_n$. Decompose \tilde{f} in the basis of irreducible representations of S_n as

$$\tilde{f}(\pi) = \sum_{\mu \in \mathcal{P}_n} \sum_{i,j=1}^{\dim(V_\mu)} \alpha_{\mu,i,j} \cdot V_\mu(\pi)_{i,j},$$

where $\alpha_{\mu,i,j} \in \mathbb{C}$. Then,

$$V_\lambda(f) = \sum_{\pi \in S_n} \sum_{\mu \in \mathcal{P}_n} \sum_{i,j=1}^{\dim(V_\mu)} \alpha_{\mu,i,j} \cdot V_\mu(\pi^{-1})_{i,j} V_\lambda(\pi).$$

Thus, (30) implies that

$$V_\lambda(f)_{k,\ell} = \frac{\alpha_{\lambda,k,\ell} n!}{\dim(V_\lambda)}.$$

We conclude that $V_\lambda(f) = 0$ for all λ with $\lambda_1 \leq n - t - 1$ if and only if $\alpha_{\mu,i,j} = 0$ for all μ with $\mu_1 \leq n - t - 1$. That is, if and only if $\tilde{f} \in W'$. Since $W = W'$ and $\tilde{f} \in W$ if and only if $f \in W$ we conclude that $V_\lambda(\tilde{f}) = 0$ for all λ with $\lambda_1 \leq n - t - 1$ if and only if $f \in W$. \square

3.4.3 Antisymmetrizers

We introduce the *column antisymmetrizer* of a tableau T ,

$$b_T := \sum_{\sigma \in Q_T} \text{sign}(\sigma) \cdot \sigma, \quad (31)$$

which is an element of the group algebra $\mathbb{C}S_n$. We study in which irreducible representations the antisymmetrizers have a non-trivial action. We start by studying this question for the Young modules since these have a simpler combinatorial nature.

Claim 3.17. *Let T be a tableau of shape $\lambda \in \mathcal{P}_n$. Then*

(i) $U_\lambda(b_T) \neq 0$.

(ii) *If $U_\mu(b_T) \neq 0$ then $\lambda \supseteq \mu$.*

Proof. We first recall the basic properties. For tabloids $[T'], [T'']$ of shape μ and $\pi \in S_n$ we have

$$U_\mu(\pi)_{[T'], [T'']} = 1_{\pi([T']) = [T']}.$$

We first establish (i). We have

$$U_\lambda(b_T)_{[T], [T]} = \sum_{\pi \in Q_T} \text{sign}(\pi) 1_{\pi([T]) = [T]} = 1,$$

since any permutation in Q_T except the identity maps $[T]$ to a different tabloid. In particular $U_\lambda(b_T) \neq 0$.

We next establish (ii). To do so, we show that for every tabloid $[S]$ of shape μ we have $b_T[S] = 0$. Assume for a moment that there exist two elements a, b in the same row of S and the same column of T . Define $s_{a,b} = \frac{1}{2}(\text{Id} - (a, b)) \in \mathbb{C}S_n$ where Id is the identity permutation and (a, b) is the permutation that swaps a and b . On the one hand $b_T * s_{a,b} = b_T$, and on the other hand $s_{a,b}[S] = 0$. Hence

$$b_T[S] = (b_T * s_{a,b})[S] = b_T(s_{a,b}[S]) = 0.$$

So, if $U_\mu(b_T) \neq 0$ such a pair cannot exist. Thus, all the elements in the first column of T appear in different rows in S ; all the elements in the second column of T appear in different rows of S , etc. This implies that $\mu' \supseteq \lambda'$ where μ', λ' are the conjugate partitions to μ, λ . This in turn implies that $\lambda \supseteq \mu$ as we claimed. \square

We now derive the analogous claim for the irreducible representations.

Claim 3.18. *Let T be a tableau of shape $\lambda \in \mathcal{P}_n$. Then*

(i) $V_\lambda(b_T) \neq 0$.

(ii) If $V_\mu(b_T) \neq 0$ then $\lambda \supseteq \mu$.

Proof. Let $\mu \in \mathcal{P}_n$ be such that $\lambda \not\supseteq \mu$. $K_{\mu,\mu} = 1$ and hence V_μ appears in the decomposition of U_μ into irreducible representations. Thus, the fact that $U_\mu(b_T) = 0$ by Claim 3.17 implies that $V_\mu(b_T) = 0$, proving (ii).

Now, since $K_{\lambda,\lambda} = 1$ and $K_{\tau,\lambda} = 0$ unless $\tau \supseteq \lambda$ we have that U_λ decomposes as the sum of V_λ plus other irreducible representations V_τ with $\tau \supset \lambda$. Since $U_\lambda(b_T) \neq 0$ by Claim 3.17 and $V_\tau(b_T) = 0$ when $\tau \supset \lambda$ by part (ii) we deduce that $V_\lambda(b_T) \neq 0$, proving (i). \square

3.4.4 Spanning vectors for W^\perp

We will prove the following lemma in this subsection.

Lemma 3.19. *Let $f \in \mathbb{C}S_n$. Then $f \in W$ iff $b_T * f = 0$ for all tableaux T of shape λ with $\lambda_1 = n - t - 1$.*

We first show that this gives a basis of integer vectors for W^\perp of small ℓ_1 norm.

Corollary 3.20. *W^\perp has a $(t+2)!$ -bounded integer basis in ℓ_1 .*

Proof. Let T be a tableau of shape $\lambda \in \mathcal{P}_n$ with $\lambda_1 = n - t - 1$. The condition $b_T * f = 0$ is equivalent to

$$\sum_{\sigma \in Q_T} \text{sign}(\sigma) f(\sigma^{-1}\pi) = 0 \quad \text{for all } \pi \in S_n.$$

The ℓ_1 norm of the vectors in W^\perp these define is $|Q_T|$, which we next derive a bound on. Let $\lambda' = (\lambda'_1, \dots, \lambda'_m)$ be the conjugate partition to λ . Then $|Q_T| = \prod_{i=1}^m \lambda'_i!$. Observe that for any $a, b \geq 0$, $b+1 = \binom{b+1}{b} \leq \binom{a+b+1}{b}$ and hence $(a+1)!(b+1)! \leq (a+b+1)!$. Thus

$$|Q_T| = \prod_{i=1}^m \lambda'_i! = \prod_{i=1}^m (\lambda'_i - 1 + 1)! \leq \left(1 + \sum_{i=1}^m (\lambda'_i - 1)\right)! = (1 + n - \lambda_1)! = (t+2)!. \quad \square$$

For the proof of Lemma 3.19 we need the following auxiliary claims.

Claim 3.21. *Let $\mu \in \mathcal{P}_n$ satisfy $\mu_1 \leq n - t - 1$. Then for every tableau S of shape μ there exists a shape $\lambda \in \mathcal{P}_n$ with $\lambda_1 = n - t - 1$, a tableau T of shape λ and an element $g \in \mathbb{C}S_n$ such that $b_S = g * b_T$.*

Proof. Define an element in a tableau as *maximal* if it is the last element in its row and its column (in English notation, it is the rightmost element in its row and the bottom element in its column). We construct T from S by iteratively moving maximal elements which are not in the first row to the end of the first row, until the first row of T contains exactly $n - t - 1$ elements. It is simple to verify that this process guarantees that Q_T is a subgroup of Q_S , since each column of T is contained in a column of S . Let $\{\sigma_1, \dots, \sigma_r\}$ be representatives for the left cosets of Q_T in Q_S . Then we have

$$b_S = \left(\sum_{i=1}^r \text{sign}(\sigma_i) \cdot \sigma_i \right) * b_T. \quad \square$$

Observe that if T, S are tableaux of shape λ and $\sigma \in S_n$ satisfies $\sigma(T) = S$, then $b_S = \sigma b_T \sigma^{-1}$. Define, for $\lambda \in \mathcal{P}_n$,

$$c_\lambda := \sum_{S \text{ of shape } \lambda} b_S = \sum_{\sigma \in S_n} \sigma b_T \sigma^{-1} \quad (32)$$

where T is an arbitrary tableau of shape λ . Denoting by Tr the trace of a matrix, the definition implies that

$$\text{Tr}(V_\mu(c_\lambda)) = n! \text{Tr}(V_\mu(b_T)) \quad \text{for } \mu \in \mathcal{P}_n. \quad (33)$$

Claim 3.22. $V_\mu(c_\lambda)$ is a multiple of the identity for all $\lambda, \mu \in \mathcal{P}_n$. Moreover, if there exists a tableau T of shape λ such that $V_\mu(b_T) \neq 0$ then $V_\mu(c_\lambda) \neq 0$.

Proof. Fix $\lambda, \mu \in \mathcal{P}_n$. By (32) we have for any $\pi \in S_n$,

$$\pi c_\lambda = \sum_{\sigma \in S_n} \pi \sigma b_T \sigma^{-1} = \sum_{\sigma \in S_n} \sigma b_t \sigma^{-1} \pi = c_\lambda \pi.$$

Hence Schur's lemma implies that $V_\mu(c_\lambda)$ is a multiple of the identity.

Now suppose that $V_\mu(b_T) \neq 0$ for some tableau T of shape λ . Observe that $b_T * b_T = |Q_T| b_T$. Thus $|Q_T|^{-1} V_\mu(b_T)$ is a projection matrix. Hence, $V_\mu(b_T) \neq 0$ implies that $\text{Tr}(V_\mu(b_T)) \neq 0$. Thus $\text{Tr}(V_\mu(c_\lambda)) \neq 0$ by (33), from which $V_\mu(c_\lambda) \neq 0$ follows. \square

Proof of Lemma 3.19. Suppose first that $f \in W$ and let T be a tableau of shape λ for some $\lambda \in \mathcal{P}_n$ satisfying $\lambda_1 = n - t - 1$. We will show that $b_T * f = 0$ by showing that $V_\mu(b_T * f) = 0$ for all $\mu \in \mathcal{P}_n$. Fix $\mu \in \mathcal{P}_n$. If $\mu_1 \leq n - t - 1$ we have $V_\mu(f) = 0$ by Claim 3.16. If $\mu_1 \geq n - t$ we have $V_\mu(b_T) = 0$ by Claim 3.18(ii). Thus in all cases $V_\mu(b_T * f) = V_\mu(b_T) V_\mu(f) = 0$.

Now suppose that $f \notin W$. By Claim 3.16 there exists some $\mu \in \mathcal{P}_n$ with $\mu_1 \leq n - t - 1$ such that $V_\mu(f) \neq 0$. Putting together Claim 3.18(i) and Claim 3.22 we have that $V_\mu(c_\mu)$ is a non-zero multiple of the identity. Thus $V_\mu(c_\mu * f) = V_\mu(c_\mu) V_\mu(f) \neq 0$. By the definition (32) of c_μ , this implies that there exists some tableau S of shape μ such that $V_\mu(b_S * f) \neq 0$. By Claim 3.21, there exists a shape $\lambda \in \mathcal{P}_n$ with $\lambda_1 = n - t - 1$, a tableau T of shape λ and an element $g \in \mathbb{C} S_n$ such that $b_S = g * b_T$. Since $V_\mu(g) V_\mu(b_T * f) = V_\mu(b_S * f) \neq 0$, we conclude that $b_T * f \neq 0$, as required. \square

3.5 The number of t -wise permutations

One may use Theorem 2.5 to estimate the number of t -wise permutations of a given size, as we did for orthogonal arrays and t -designs. To this end, one needs to calculate the parameter $\rho(V)$ defined by (13). We leave this calculation for future work but present in this section the results of numerical calculations which give some evidence that $\rho(V)$ has a nice product structure.

Let $B = S_n$ and A be as in (28). Let ϕ be the $B \times A$ matrix whose columns are given by the (f_σ) , $\sigma \in A$, defined in (29). As proven in Section 3.4.2, the columns of ϕ form a basis for W and $\mathcal{L}(\phi) = \mathbb{Z}^A$. Thus

$$\rho(V) = \frac{\det(\mathcal{L}(\phi))}{\sqrt{\det(\phi^t \phi)}} = \frac{1}{\sqrt{\det(\phi^t \phi)}}$$

for this matrix ϕ . Below we present the results of numerical calculations of $\det(\phi^t \phi)$ for a few small values of n and t .

$n \setminus t$	1	2	3
3	$3 \cdot 2$	1	—
4	$3 \cdot 2^{18}$	$3 \cdot 2^3$	1
5	$5^9 3^{17} 2^{19}$	$5^9 3^2 2^{84}$	$5 \cdot 3 \cdot 2^3$
6	$5 \cdot 3^{42} 2^{94}$	$5^{64} 3^{162} 2^{276}$?
7	$7^{25} 5^{37} 3^{38} 2^{112}$	$7^{100} 5^{65} 3^{627} 2^{1150}$?
8	$7 \cdot 5^{50} 3^{100} 2^{308}$?	?

4 Proof of main theorems

We prove our main theorems, Theorem 2.4 and Theorem 2.5, in this section. We start by stating a local central limit theorem from which our main theorems will follow.

4.1 Local central limit theorem statement

Let B be a finite set and V be a vector space of functions from B to the rational numbers \mathbb{Q} . Let $\{\phi_a : B \rightarrow \mathbb{Q}\}_{a \in A}$ be a basis for V , where A is some finite index set of size $\dim(V)$. This basis is arbitrary for now but will be chosen in a convenient way in the next subsection. Let $\phi : B \rightarrow \mathbb{Q}^A$ be defined as $\phi(b)_a = \phi_a(b)$. It may be useful to think of ϕ as a $B \times A$ matrix, whose entries are $\phi_a(b)$. Fix $0 < p < 1$ and define T to be a random subset of B , with each point of B chosen independently into T with probability p . In other words, we let $\{T_b\}$, $b \in B$, be a collection of independent identically distributed random variables with $\mathbb{P}[T_b = 1] = 1 - \mathbb{P}[T_b = 0] = p$ and let $T := \{b \in B : T_b = 1\}$. Define

$$X := \sum_{b \in B} T_b \cdot \phi(b) \in \mathcal{L}(\phi), \quad (34)$$

where $\mathcal{L}(\phi)$ is the lattice in \mathbb{Q}^A generated by $\{\phi(b)\}$, $b \in B$. Our main Theorems will follow from a precise estimate of the probability $\mathbb{P}[X = \mathbb{E}[X]]$. Along the way, however, we will pass through estimating $\mathbb{P}[X = \lambda]$ for an arbitrary point $\lambda \in \mathcal{L}(\phi)$ (though our estimate will only be meaningful for λ close to $\mathbb{E}[X]$). Since this is a useful result in itself, which also requires less assumptions, we encapsulate it in the following theorem. We note that the mean of X is given by

$$\mathbb{E}[X] = p \sum_{b \in B} \phi(b)$$

and the covariance matrix of X is given by

$$\Sigma[X] := \mathbb{E}[(X - \mathbb{E}[X])^t (X - \mathbb{E}[X])] = p(1-p)\phi^t \phi \quad (35)$$

where $\phi^t \phi$ is the symmetric positive definite $A \times A$ matrix satisfying $(\phi^t \phi)_{a,a'} = \sum_{b \in B} \phi(b)_a \phi(b)_{a'}$. The positive definite property follows from the fact that the $\{\phi_a\}$, $a \in A$, are linearly independent.

Theorem 4.1 (Local central limit theorem). *There exists a constant $C > 0$ such that the following is true. Assume that the following conditions hold for some integers $c_2, c_3 \geq 1$,*

1. *Boundedness of V : V has a c_2 -bounded integer basis in ℓ_∞ .*
2. *Boundedness of V^\perp : V^\perp has a c_3 -bounded integer basis in ℓ_1 .*
3. *Symmetry: for any $b_1, b_2 \in B$ there exists a symmetry π of V satisfying $\pi(b_1) = b_2$.*

If

$$\min(p|B|, (1-p)|B|) \geq C \cdot c_2 c_3^2 \dim(V)^6 \log(2c_3 \dim(V))^6$$

then for every $\lambda \in \mathcal{L}(\phi)$,

$$\mathbb{P}[X = \lambda] = \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{\dim(V)}{2}} \sqrt{\det \Sigma[X]}} \left(e^{-\frac{1}{2}(\lambda - \mathbb{E}[X])^t \Sigma[X]^{-1} (\lambda - \mathbb{E}[X])} + \delta(\lambda) \right) \quad (36)$$

$$\text{with } |\delta(\lambda)| \leq \frac{C \dim(V)^3 (\log(2c_2 \dim(V)))^{3/2}}{\sqrt{\min(p|B|, (1-p)|B|)}}.$$

We point out explicitly that this theorem does not require the divisibility or constant functions assumptions of our main theorems, Theorem 2.4 and Theorem 2.5. In Subsection 4.7 below we explain how our main theorems follow (easily) from this local central limit theorem and the extra assumptions by applying the theorem with $\lambda = \mathbb{E}[X]$.

We also note that the local central limit theorem does not depend on our choice of basis ϕ in the sense that if it holds for one basis it holds for all bases. We make this fact more explicit in Subsection 4.8 where we state an equivalent basis-free version of the theorem.

The local central limit theorem is proved in Subsections 4.2 to 4.6 below.

4.2 Fourier analysis

We continue with the notation of the previous section and assume the conditions of Theorem 4.1. We fix $\{\phi_a\}$, $a \in A$ to be the basis of integer-valued functions, satisfying $\|\phi_a\|_\infty \leq c_2$ for all $a \in A$, whose existence is guaranteed by the boundedness condition for V . We also make the simplifying assumption

$$p \leq \frac{1}{2}.$$

Near the end of the proof we will show how to get rid of this assumption by utilizing the bijection $T \mapsto B \setminus T$. Finally, we denote

$$N := p|B|.$$

We stress that N need not be an integer in our proof of the local central limit theorem. However, when we later deduce our main theorems from the local central limit theorem, we will choose p in such a way that N will be an integer.

Our main technique to study the distribution of X is Fourier analysis. The Fourier transform of X is the function $\hat{X} : \mathbb{R}^A \rightarrow \mathbb{C}$ defined by

$$\hat{X}(\theta) := \mathbb{E}[e^{2\pi i \langle X, \theta \rangle}],$$

where $\langle X, \theta \rangle := \sum_{a \in A} X_a \theta_a$. Define the dual lattice L to $\mathcal{L}(\phi)$ (sometimes called the annihilator or reciprocal lattice of $\mathcal{L}(\phi)$), as the set of vectors in \mathbb{R}^A having an integer inner product with the vectors of $\mathcal{L}(\phi)$. That is,

$$L := \{\theta \in \mathbb{R}^A : \langle \theta, \lambda \rangle \in \mathbb{Z} \quad \forall \lambda \in \mathcal{L}(\phi)\}.$$

Noting that $\mathcal{L}(\phi)$ has full rank, since the $\{\phi_a\}$, $a \in A$, are linearly independent, it follows that L is also a full rank lattice and the relation $\det(\mathcal{L}(\phi)) \det(L) = 1$ holds. Since $e^{2\pi i \langle X, \alpha \rangle} = 1$ almost surely when $\alpha \in L$, we see that \hat{X} is L -periodic,

$$\hat{X}(\theta + \alpha) = \hat{X}(\theta) \quad \forall \theta \in \mathbb{R}^A, \alpha \in L. \tag{37}$$

The covariance matrix of X provides a natural norm to work with in Fourier space. Define

$$R := \phi^t \phi,$$

so that $\Sigma[X] = p(1-p)R$ by (35). As mentioned, R is a symmetric positive definite $A \times A$ matrix satisfying $R_{a,a'} = \sum_{b \in B} \phi(b)_a \phi(b)_{a'}$. We define a norm in Fourier space by

$$\|\theta\|_R := \left(\frac{1}{|B|} \theta^t R \theta \right)^{1/2} = \left(\frac{1}{|B|} \sum_{b \in B} \langle \phi(b), \theta \rangle^2 \right)^{1/2} \quad (\theta \in \mathbb{R}^A).$$

Balls in the R -norm are denoted by

$$\mathcal{B}_R(\varepsilon) := \{\theta \in \mathbb{R}^A : \|\theta\|_R \leq \varepsilon\}.$$

Let D be the Voronoi cell of 0 in the lattice L , with respect to the R -norm. That is,

$$D := \{\theta \in \mathbb{R}^A : \|\theta\|_R < \|\theta - \alpha\|_R \quad \forall \alpha \in L \setminus \{0\}\}. \quad (38)$$

Observe that D is a bounded set since L has full rank. Moreover, $\alpha + D$ and $\alpha' + D$ are disjoint for distinct $\alpha, \alpha' \in L$, and $\cup_{\alpha \in L} (\alpha + D)$ covers all of \mathbb{R}^A except a set of Lebesgue measure zero (since only a Lebesgue measure zero of points in \mathbb{R}^A are equidistant to two points in L). It follows that $\text{Vol}(D) = \det(L) = \det(\mathcal{L}(\phi))^{-1}$, where Vol denotes Lebesgue measure, and that we have the following inversion formula.

Fact 4.2 (Fourier inversion formula on lattices).

$$\mathbb{P}[X = \lambda] = \det(\mathcal{L}(\phi)) \int_D \widehat{X}(\theta) e^{-2\pi i \langle \lambda, \theta \rangle} d\theta \quad \forall \lambda \in \mathcal{L}(\phi).$$

Thus, our goal from now on is to understand the Fourier transform of X . We start with an explicit formula for \widehat{X} .

Claim 4.3. *We have*

$$\widehat{X}(\theta) = \prod_{b \in B} \left(1 - p + p e^{2\pi i \langle \phi(b), \theta \rangle}\right).$$

Proof. By definition $X = \sum_{b \in B} T_b \phi(b)$, where $T_b \in \{0, 1\}$ are independent with $\mathbb{P}[T_b = 1] = p$. Thus

$$\begin{aligned} \widehat{X}(\theta) &= \mathbb{E}[e^{2\pi i \langle X, \theta \rangle}] = \mathbb{E}[e^{2\pi i \sum_{b \in B} T_b \langle \phi(b), \theta \rangle}] \\ &= \prod_{b \in B} \mathbb{E}[e^{2\pi i T_b \langle \phi(b), \theta \rangle}] = \prod_{b \in B} (1 - p + p e^{2\pi i \langle \phi(b), \theta \rangle}). \end{aligned} \quad \square$$

An important ingredient in controlling \widehat{X} is the following property. The terms $\langle \phi(b), \theta \rangle$ which arise in the Fourier transform $\widehat{X}(\theta)$ are tame in the following sense: if most of them are small, then all of them are small; and if most of them are close to integers, then all of them are close to integers. This is captured by the following lemma. The constant c_2 appearing in the lemma is the one given in the boundedness assumption for V .

Lemma 4.4. *There exists a universal constant $C > 0$ such that if we set*

$$M := C(|A| \log(2c_2|A|))^{3/2} \quad (39)$$

then for every $\theta \in \mathbb{R}^A$:

1.

$$\max_{b \in B} |\langle \phi(b), \theta \rangle| \leq M \left(\frac{1}{|B|} \sum_{b \in B} \langle \phi(b), \theta \rangle^2 \right)^{1/2} = M \|\theta\|_R.$$

2. Write $\langle \phi(b), \theta \rangle = n_b + r_b$, where $n_b \in \mathbb{Z}$ and $r_b \in [-1/2, 1/2)$. Then

$$\max_{b \in B} |r_b| \leq M \left(\frac{1}{|B|} \sum_{b \in B} r_b^2 \right)^{1/2}.$$

We note that the proof of the lemma uses only the boundedness assumption for V and the assumption that V has a transitive symmetry group and it is the only place where the symmetry assumption is used. We prove Lemma 4.4 in Subsection 4.3. The main ingredient in its proof is the notion of local correctability of the map ϕ .

Our next step is to approximate \hat{X} near zero. The next lemma achieves this by approximating $\hat{X}(\theta)$ by its Taylor expansion at zero for $\theta \in \mathcal{B}_R(\varepsilon)$.

Lemma 4.5 (Estimating the Fourier transform near zero). *For all $0 < \varepsilon \leq \frac{1}{8M}$ and $\theta \in \mathcal{B}_R(\varepsilon)$,*

$$\hat{X}(\theta) = \exp(2\pi i \cdot \langle \mathbb{E}[X], \theta \rangle - 2\pi^2 \cdot \theta^t \Sigma[X] \theta + \delta(\theta))$$

where $|\delta(\theta)| = O(M \|\theta\|_R^3 N)$.

In this lemma as well as in the remainder of the paper we use the $O(\cdot)$ notation to hide universal constants, independent of all other parameters. We prove Lemma 4.5 in Subsection 4.4. We next derive an upper bound on the Fourier transform at points which are far from zero. Recalling that $\hat{X}(\theta)$ is an L -periodic function, such a bound can only hold for θ bounded away from the points of L . We achieve this by requiring θ to belong to $D \setminus \mathcal{B}_R(\varepsilon)$.

Lemma 4.6 (Bounding the Fourier transform far from L). *For all $\varepsilon > 0$ and $\theta \in D \setminus \mathcal{B}_R(\varepsilon)$,*

$$|\hat{X}(\theta)| \leq \exp(-\beta^2 N)$$

where $\beta = \beta(\varepsilon) = \min(\varepsilon, \frac{1}{c_3 M})$.

We prove Lemma 4.6 in Subsection 4.5. This lemma is the only place where we use the boundedness assumption for V^\perp .

4.3 Local correction

A map $\psi : B \rightarrow \mathbb{Z}^A$ is said to be locally correctable, if for any small subset $E \subset B$ and any $e \in E$, we can express $\psi(e)$ as a short integer combination of $\{\psi(b) : b \in B \setminus E\}$. This is an analog of the local correction property of codes, usually studied over finite fields.

Definition 4.7 (Locally correctable). *A map $\psi : B \rightarrow \mathbb{Z}^A$ is called (δ, s) -locally correctable if for any $E \subset B$ of size $1 \leq |E| \leq \delta|B|$ and any $e \in E$, there exists a $\gamma \in \mathbb{Z}^{B \setminus E}$ with $\|\gamma\|_1 \leq s$ such that*

$$\psi(e) = \sum_{b \in B \setminus E} \gamma_b \cdot \psi(b).$$

Regarding ψ as a $B \times A$ matrix, we see that local correctability is actually a property of the space W spanned by the columns of ψ and does not depend on the particular choice of basis given by ψ . Still, it is convenient to define local correctability this way since our usage for it will be with a particular choice of basis. We say that a permutation $\pi \in S_B$ is a *symmetry* of ψ if it is a symmetry of W . We show that the assumptions that ψ is bounded and has a transitive symmetry group imply that it is locally correctable. We use the notation $\|\psi\|_\infty := \max_{a \in A, b \in B} |\psi(b)_a|$, $\psi(\gamma) := \sum_{b \in B} \gamma_b \psi(b)$ for $\gamma \in \mathbb{R}^B$ and $\psi(S) = \sum_{b \in S} \psi(b)$ for $S \subset B$.

Lemma 4.8. *Let $\psi : B \rightarrow \mathbb{Z}^A$ be such that $\|\psi\|_\infty \leq c$ and such that the symmetry group of ψ acts transitively on B . Then ψ is (δ, s) -locally correctable for some $s = O(|A| \log(2c|A|))$ and $\delta = \frac{1}{8s}$.*

We need the following auxiliary claim.

Claim 4.9. *Let $\psi : B \rightarrow \mathbb{Z}^A$ be such that $\|\psi\|_\infty \leq c$. Then for any subset $S \subset B$ of size $|S| \geq O(|A| \log(2c|A|))$ there exists a vector $\gamma \in \{-1, 0, 1\}^S$ having at least $\frac{1}{4}|S|$ non-zero coordinates and satisfying $\psi(\gamma) = 0$.*

Proof. The claim follows from the pigeon hole principle. Let $\alpha \in (0, 1)$ be such that for any $n \geq 1$, the number of strings in $\{0, 1\}^n$ having less than $\frac{1}{4}n$ ones is at most $2^{\alpha n}$. Fix $S \subset B$. For a subset $S' \subseteq S$, we have that $\psi(S')$ is an $|A|$ -dimensional integer vector with entries bounded by $c|S|$ in absolute value and hence the total number of distinct values for it is bounded by $(2c|S| + 1)^{|A|}$. The number of subsets of S is $2^{|S|}$. Hence, if

$$2^{(1-\alpha)|S|} > (2c|S| + 1)^{|A|} \quad (40)$$

there must exist two distinct subsets S_1, S_2 such that $\psi(S_1) = \psi(S_2)$ and $|S_1 \triangle S_2| \geq \frac{1}{4}|S|$ (where we use \triangle to denote symmetric difference). We then set $\gamma = 1_{S_1} - 1_{S_2}$ and have that $\psi(\gamma) = 0$ and γ has at least $\frac{1}{4}|S|$ non-zero coordinates. Now, it is a simple exercise to verify that the condition $|S| \geq O(|A| \log(2c|A|))$ with a large enough hidden constant implies (40). \square

Proof of Lemma 4.8. Let $s = O(|A| \log(2c|A|))$ be an integer larger than the lower bound on $|S|$ given by Claim 4.9. Set $\delta = \frac{1}{8s}$. Assume that $\delta|B| \geq 1$ since otherwise the claim holds vacuously. Fix a subset $E \subset B$ of size $1 \leq |E| \leq \delta|B|$ and an element $e \in E$. For each $b \in B$, let π_b be a symmetry of ψ satisfying $\pi_b(b) = e$.

Choose S uniformly among the subsets of size s of B and choose b_0 uniformly in S . Let $\gamma \in \{-1, 0, 1\}^S$ be some (random) vector having at least $\frac{1}{4}|S|$ non-zero coordinates and satisfying $\psi(\gamma) = 0$, whose existence is guaranteed by Claim 4.9. Define the events

$$\begin{aligned} \Omega_1 &:= \{\gamma_{b_0} = 0\}, \\ \Omega_2 &:= \{\exists b \in S \setminus \{b_0\} \text{ satisfying } \pi_{b_0}(b) \in E\}. \end{aligned}$$

We first show that $\Omega_1^c \cap \Omega_2^c$ has positive probability. Indeed, this follows from the fact that $\mathbb{P}[\Omega_1] \leq \frac{3}{4}$ by the properties of γ and, observing that conditioned on b_0 each $b \in S \setminus \{b_0\}$ is uniformly distributed in $B \setminus \{b_0\}$,

$$\mathbb{P}[\Omega_2] \leq s \frac{|E|}{|B|} \leq s\delta = \frac{1}{8}.$$

Now, assume that $\Omega_1^c \cap \Omega_2^c$ occurred. We may assume WLOG that $\gamma_{b_0} = 1$ (since otherwise we may replace γ by $-\gamma$). Thus

$$\psi(b_0) = - \sum_{b \in S \setminus \{b_0\}} \gamma_b \psi(b). \quad (41)$$

Since π_{b_0} is a symmetry of ψ , there exists an invertible linear map $\tau : \mathbb{Q}^A \rightarrow \mathbb{Q}^A$ such that $\psi(\pi_{b_0}(b)) = \tau(\psi(b))$ for all $b \in B$. Applying τ to (41) gives

$$\psi(e) = \tau(\psi(b_0)) = - \sum_{b \in S \setminus \{b_0\}} \gamma_b \tau(\psi(b)) = - \sum_{b \in S \setminus \{b_0\}} \gamma_b \psi(\pi_{b_0}(b)).$$

Finally, Ω_2^c implies that $\pi_{b_0}(b) \notin E$ for all $b \in S \setminus \{b_0\}$ and the lemma follows. \square

We now derive Lemma 4.4.

Proof of Lemma 4.4. We have by Lemma 4.8 that ϕ is (δ, s) -locally correctable with $s = O(|A| \log(2c_2|A|))$ and $\delta = \frac{1}{8s}$. We will establish Lemma 4.4 with $M = s/\sqrt{\delta}$.

Let us first prove the first item. Let $\beta := \left(\frac{1}{|B|} \sum_{b \in B} |\langle \phi(b), \theta \rangle|^2 \right)^{1/2}$ and set $E := \{b \in B : |\langle \phi(b), \theta \rangle| \geq \frac{\beta}{\sqrt{\delta}}\}$. Then we must have that $|E| \leq \delta|B|$. If E is empty we are done. Otherwise, since ϕ is (δ, s) -locally correctable, for any $e \in E$ we can express $\phi(e)$ as $\phi(e) = \sum_{b \notin E} \gamma_b \cdot \phi(b)$ where $\sum |\gamma_b| \leq s$. Hence in particular,

$$|\langle \phi(e), \theta \rangle| = \left| \sum_{b \notin E} \gamma_b \langle \phi(b), \theta \rangle \right| \leq \sum_{b \notin E} |\gamma_b| \cdot |\langle \phi(b), \theta \rangle| \leq s \cdot \frac{\beta}{\sqrt{\delta}}.$$

The second item is very similar. Let $\beta := \left(\frac{1}{|B|} \sum_{b \in B} |r_b|^2 \right)^{1/2}$ and set $E := \{b \in B : |r_b| \geq \frac{\beta}{\sqrt{\delta}}\}$. Again, $|E| \leq \delta|B|$ and thus any $e \in E$ can be expressed as $\phi(e) = \sum_{b \notin E} \gamma_b \cdot \phi(b)$ with $\|\gamma\|_1 \leq s$. Hence

$$\langle \phi(e), \theta \rangle = \sum_{b \notin E} \gamma_b \langle \phi(b), \theta \rangle = \sum_{b \notin E} \gamma_b (n_b + r_b),$$

and in particular $r_e = \sum_{b \notin E} \gamma_b r_b \pmod{1}$ (where we mean that the modulo 1 maps \mathbb{R} to $[-1/2, 1/2)$). Hence

$$|r_e| \leq \sum_{b \notin E} |\gamma_b| |r_b| \leq s \cdot \frac{\beta}{\sqrt{\delta}}. \quad \square$$

4.4 Estimating the Fourier transform near zero

We prove Lemma 4.5 in this subsection. Let $\theta \in \mathcal{B}_R(\varepsilon) \subset \mathbb{R}^A$. Recall that by Claim 4.3 we have that

$$\widehat{X}(\theta) = \prod_{b \in B} \left(1 - p + p \cdot e^{2\pi i \langle \phi(b), \theta \rangle} \right).$$

Let us shorthand $x_b = 2\pi \langle \phi(b), \theta \rangle$. By our assumptions that $\theta \in \mathcal{B}_R(\varepsilon)$ and Lemma 4.4 we have

$$\max_{b \in B} |x_b| = 2\pi \max_{b \in B} |\langle \phi(b), \theta \rangle| \leq 2\pi M \|\theta\|_R \leq \frac{\pi}{4}, \quad (42)$$

where the last inequality follows from the assumption that $\|\theta\|_R \leq \varepsilon \leq \frac{1}{8M}$. Define the function $f : \mathbb{R} \rightarrow \mathbb{C}$ given by $f(x) := 1 - p + pe^{ix}$. Then

$$\widehat{X}(\theta) = \prod_{b \in B} f(x_b). \quad (43)$$

Claim 4.10. *If $0 \leq p \leq 1$ and $|x| \leq \frac{\pi}{4}$ then $f(x) = \exp(ipx - \frac{1}{2}p(1-p)x^2 + \delta(x))$ where $|\delta(x)| = O(p|x|^3)$.*

Proof. Let $y = p(e^{ix} - 1)$ so that $f(x) = 1 + y$. Our assumptions imply that $|y| \leq |e^{ix} - 1| \leq \sqrt{2 - \sqrt{2}} < 1$ so that $\log(1 + y) = y - y^2/2 + O(|y|^3)$. Now, $y = ipx - px^2/2 + O(p|x|^3)$ and $y^2 = -p^2x^2 + O(p^2|x|^3)$. Hence

$$\log(f(x)) = ipx - \frac{1}{2}p(1-p)x^2 + O(p|x|^3). \quad \square$$

Applying Claim 4.10 to each term in (43) we obtain

$$\begin{aligned}\widehat{X}(\theta) &= \exp(2\pi i \cdot p \cdot \sum_{b \in B} \langle \phi(b), \theta \rangle - 4\pi^2 \cdot \frac{1}{2} p(1-p) \cdot \sum_{b \in B} \langle \phi(b), \theta \rangle^2 + \delta(\theta)) \\ &= \exp(2\pi i \cdot \langle \mathbb{E}[X], \theta \rangle - 2\pi^2 \cdot \theta^t \Sigma[X] \theta + \delta(\theta))\end{aligned}\tag{44}$$

where we recall that $\mathbb{E}[X] = p \sum_{b \in B} \phi(b)$ and that $\Sigma[X]_{a,a'} = p(1-p) \sum_{b \in B} \phi(b)_a \phi(b)_{a'}$, for $a, a' \in A$. The error term is bounded by

$$|\delta(\theta)| = O(p \sum_{b \in B} |\langle \phi(b), \theta \rangle|^3).\tag{45}$$

Applying (42) again yields

$$\sum_{b \in B} |\langle \phi(b), \theta \rangle|^3 \leq \max_{b \in B} |\langle \phi(b), \theta \rangle| \cdot \sum_{b \in B} |\langle \phi(b), \theta \rangle|^2 \leq M \|\theta\|_R \cdot \|\theta\|_R^2 |B| = M \|\theta\|_R^3 |B|,$$

and since $p|B| = N$ we can bound the error term by

$$|\delta(\theta)| \leq O(M \|\theta\|_R^3 N).$$

4.5 Bounding the Fourier transform far from L

We next bound the Fourier transform far from 0 in the R -norm, proving Lemma 4.6. Fix $\varepsilon > 0$ and $\theta \in D \setminus \mathcal{B}_R(\varepsilon)$. Our goal is to show that $|\widehat{X}(\theta)|$ must be small. Let us decompose $\langle \phi(b), \theta \rangle = n_b + r_b$ where $n_b \in \mathbb{Z}$ and $r_b \in [-1/2, 1/2]$. Recall that by Claim 4.3 we have that

$$\widehat{X}(\theta) = \prod_{b \in B} (1 - p + p \cdot e^{2\pi i \langle \phi(b), \theta \rangle}) = \prod_{b \in B} (1 - p + p \cdot e^{2\pi i r_b}).\tag{46}$$

We need two auxiliary claims.

Claim 4.11. $|\widehat{X}(\theta)| \leq \exp(-\frac{N}{|B|} \sum_{b \in B} r_b^2) \leq \exp(-\frac{N}{M^2} \max_{b \in B} r_b^2)$, where M is defined in (39).

Proof. It is simple to verify that for any $|x| \leq 1/2$ and $p \leq 1/2$,

$$|1 - p + p e^{2\pi i x}| \leq \exp(-p x^2).$$

Hence

$$|\widehat{X}(\theta)| \leq \exp(-p \sum_{b \in B} r_b^2) = \exp(-\frac{N}{|B|} \sum_{b \in B} r_b^2).$$

The second inequality now follows from the second part of Lemma 4.4. \square

Recall that c_3 is the constant from our assumption that V^\perp has a bounded integer basis in ℓ_1 . We next argue that if $\max_{b \in B} |r_b| < \frac{1}{c_3}$ then the vector $(n_b)_{b \in B}$ belongs to the space V .

Claim 4.12. *If $\max_{b \in B} |r_b| < 1/c_3$ then there exists $\alpha \in \mathbb{Q}^A$ such that $\langle \phi(b), \alpha \rangle = n_b$ for all $b \in B$.*

We remark that this is the only place in our proof where the assumption that V^\perp has a bounded integer basis in ℓ_1 is used.

Proof of Claim 4.12. Assume to the contrary that no such α exists. Then $(n_b)_{b \in B}$ does not belong to V and hence it must violate some constraint of V^\perp . However, by assumption, V^\perp is spanned by integer vectors of ℓ_1 norm at most c_3 . Hence, there exists $\gamma \in \mathbb{Z}^B$, $\|\gamma\|_1 \leq c_3$ such that

$$\sum_{b \in B} \gamma_b n_b \neq 0.$$

Since both γ and (n_b) are integer vectors, we must have that

$$|\sum_{b \in B} \gamma_b n_b| \geq 1.$$

However, we know that the vector $(n_b + r_b)_{b \in B}$ belongs to V (more precisely, to the span over \mathbb{R} of the vectors in V). Hence

$$\sum_{b \in B} \gamma_b (n_b + r_b) = 0.$$

Thus we conclude that

$$|\sum_{b \in B} \gamma_b r_b| \geq 1.$$

This is, however, impossible if $|r_b| < 1/c_3$ for all $b \in B$. \square

We now conclude the proof of Lemma 4.6. There are two cases to consider. Suppose first that $\max_{b \in B} |r_b| \geq \frac{1}{c_3}$. Then Claim 4.11 implies that

$$|\hat{X}(\theta)| \leq \exp\left(-\frac{N}{M^2 c_3^2}\right) \quad \text{if } \max_{b \in B} |r_b| \geq \frac{1}{c_3}. \quad (47)$$

Now assume instead that $\max_{b \in B} |r_b| < \frac{1}{c_3}$ and let $\alpha \in \mathbb{Q}^A$ be as in Claim 4.12. By definition, α belongs to the lattice L . It follows that

$$\|\theta - \alpha\|_R^2 = \frac{1}{|B|} \sum_{b \in B} \langle \phi(b), \theta - \alpha \rangle^2 = \frac{1}{|B|} \sum_{b \in B} r_b^2.$$

Thus, by the definition (38) of the Voronoi cell D and the fact that $\theta \in D$ and $\alpha \in L$, we deduce that

$$\|\theta\|_R^2 \leq \frac{1}{|B|} \sum_{b \in B} r_b^2.$$

Since $\theta \notin \mathcal{B}_R(\varepsilon)$ we also have $\|\theta\|_R > \varepsilon$ and thus Claim 4.11 shows that

$$|\hat{X}(\theta)| \leq \exp(-N\varepsilon^2) \quad \text{if } \max_{b \in B} |r_b| < \frac{1}{c_3}. \quad (48)$$

Taken together, (47) and (48) prove Lemma 4.6.

4.6 Proof of central limit theorem from auxiliary lemmas

In this section we prove Theorem 4.1. We start with a bound on the in-radius of D in the R -norm.

Claim 4.13. *If $\varepsilon < \frac{1}{2M}$ then $\mathcal{B}_R(\varepsilon) \subset D$.*

Proof. Let $0 \neq \alpha \in L$. By definition, $\langle \phi(b), \alpha \rangle \in \mathbb{Z}$ for all $b \in B$. Since $\mathcal{L}(\phi)$ is of full rank and $\alpha \neq 0$, there exists some $b \in B$ for which $|\langle \phi(b), \alpha \rangle| \geq 1$. It follows from Lemma 4.4 that $\|\alpha\|_R \geq \frac{1}{M}$. Since α is arbitrary, we deduce from the definition (38) of D that $\mathcal{B}_R(\varepsilon) \subset D$ for any $\varepsilon < \frac{1}{2M}$. \square

Now fix $\lambda \in \mathcal{L}(\phi)$ and recall from Fact 4.2 that

$$\mathbb{P}[X = \lambda] = \det(\mathcal{L}(\phi)) \int_D \widehat{X}(\theta) e^{-2\pi i \langle \lambda, \theta \rangle} d\theta \quad \forall \lambda \in \mathcal{L}(\phi). \quad (49)$$

Introduce a second random vector $Y \in \mathbb{R}^A$ having the Gaussian distribution with mean $\mathbb{E}[X]$ and covariance matrix $\Sigma[X]$ (that is, with the same mean and covariance matrix as X). Recall that the density function f_Y of Y equals

$$f_Y(x) := \frac{\exp(-\frac{1}{2}(x - \mathbb{E}[X])^t \Sigma[X]^{-1} (x - \mathbb{E}[X]))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det \Sigma[X]}}, \quad (50)$$

and that the Fourier transform of Y equals

$$\widehat{Y}(\theta) := \mathbb{E}[e^{2\pi i \langle Y, \theta \rangle}] = e^{2\pi i \langle \mathbb{E}[X], \theta \rangle - 2\pi^2 \theta^t \Sigma[X] \theta}. \quad (51)$$

Moreover, the Fourier inversion formula applied to Y yields

$$f_Y(x) = \int_{\mathbb{R}^A} \widehat{Y}(\theta) e^{-2\pi i \langle x, \theta \rangle} d\theta \quad \forall x \in \mathbb{R}^A. \quad (52)$$

Theorem 4.1 will follow by showing that $\mathbb{P}[X = \lambda]$ approximately equals $\det(\mathcal{L}(\phi)) f_Y(\lambda)$. Fix $0 < \varepsilon < \frac{1}{2M}$ whose exact value will be chosen later (see (58) and (59)). Combining (49), (52) and Claim 4.13 we may write

$$\begin{aligned} |\mathbb{P}[X = \lambda] - \det(\mathcal{L}(\phi)) f_Y(\lambda)| \leq \\ \det(\mathcal{L}(\phi)) \left(\underbrace{\int_{\mathcal{B}_R(\varepsilon)} |\widehat{X}(\theta) - \widehat{Y}(\theta)| d\theta}_{=: I_1} + \underbrace{\int_{D \setminus \mathcal{B}_R(\varepsilon)} |\widehat{X}(\theta)| d\theta}_{=: I_2} + \underbrace{\int_{\mathbb{R}^A \setminus \mathcal{B}_R(\varepsilon)} |\widehat{Y}(\theta)| d\theta}_{=: I_3} \right). \end{aligned} \quad (53)$$

Our next lemma provides upper bounds on each of the above integrals.

Lemma 4.14. *There exists a universal constant $c > 0$ such that:*

1. *If $\varepsilon \leq \frac{1}{8M}$ and $M\varepsilon^3 N \leq c$ then*

$$I_1 \leq \frac{M|A|^{3/2}}{2\sqrt{N}(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma[X])}}.$$

2. *If $\varepsilon \leq \frac{1}{c_3 M}$ then*

$$I_2 \leq \frac{e^{-\varepsilon^2 N}}{\det(\mathcal{L}(\phi))}.$$

3. *If $\varepsilon^2 N \geq \frac{2|A|}{\pi^2}$ then*

$$I_3 \leq \frac{e^{-\frac{1}{4}\pi^2 \varepsilon^2 N}}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma[X])}}.$$

Proof. We start with the second item. By Lemma 4.6, if $\varepsilon \leq \frac{1}{c_3 M}$ and $\theta \in D \setminus \mathcal{B}_R(\varepsilon)$ then $|\hat{X}(\theta)| \leq \exp(-\varepsilon^2 N)$. Hence,

$$I_2 \leq \text{Vol}(D) e^{-\varepsilon^2 N} = \frac{e^{-\varepsilon^2 N}}{\det(\mathcal{L}(\phi))}.$$

We continue with the third item. By (51) we have

$$I_3 = \int_{\mathbb{R}^A \setminus \mathcal{B}_R(\varepsilon)} e^{-2\pi^2 \theta^t \Sigma[X] \theta} d\theta. \quad (54)$$

To evaluate the integral let G be a standard multivariate Gaussian random vector in \mathbb{R}^A (with mean zero and identity covariance matrix). Recalling that $\Sigma[X]$ is a positive definite matrix, let $\Sigma[X]^{-1/2}$ be a symmetric positive definite matrix such that $(\Sigma[X]^{-1/2})^2 = \Sigma[X]^{-1}$. It follows that

$$Z := \frac{1}{2\pi} \Sigma[X]^{-1/2} G$$

has a Gaussian distribution with mean zero and covariance matrix $\frac{1}{4\pi^2} \Sigma[X]^{-1}$. Thus the density function of Z is

$$f_Z(\theta) = (2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma)} e^{-2\pi^2 \theta^t \Sigma[X] \theta}. \quad (55)$$

Comparing with (54) yields

$$I_3 = (2\pi)^{-\frac{|A|}{2}} \det(\Sigma)^{-\frac{1}{2}} \mathbb{P}[\|Z\|_R > \varepsilon] \leq (2\pi)^{-\frac{|A|}{2}} \det(\Sigma)^{-\frac{1}{2}} \mathbb{P}[\|G\|_2^2 > 2\pi^2 \varepsilon^2 N], \quad (56)$$

where in the last inequality we used that

$$\|Z\|_R^2 = \frac{1}{|B|} Z^t R Z = \frac{1}{4\pi^2 |B|} G^t \Sigma[X]^{-1/2} R \Sigma[X]^{-1/2} G = \frac{\|G\|_2^2}{4\pi^2 p(1-p)|B|} \leq \frac{\|G\|_2^2}{2\pi^2 N}, \quad (57)$$

recalling that $p|B| = N$ and our standing assumption that $p \leq \frac{1}{2}$. Now, The distribution of $\|G\|_2^2$ is chi-squared with $|A|$ degrees of freedom. Observing that $\mathbb{E} e^{t\|G\|_2^2} = (1-2t)^{-|A|/2}$ for $t < 1/2$ and fixing $t = 1/4$, Markov's inequality yields for any $\rho \geq 4|A|$ that

$$\mathbb{P}[\|G\|_2^2 > \rho] \leq \frac{\mathbb{E} e^{\|G\|_2^2/4}}{e^{\rho/4}} = 2^{|A|/2} e^{-\rho/4} \leq e^{-\rho/8}.$$

Applying this result to (56) and using our assumption that $\varepsilon^2 N \geq \frac{2|A|}{\pi^2}$ we have

$$I_3 \leq \frac{e^{-\frac{1}{4}\pi^2 \varepsilon^2 N}}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma)}}.$$

Lastly, we verify the first item. Using Lemma 4.5 and (51) we have

$$I_1 = \int_{\mathcal{B}_R(\varepsilon)} e^{-2\pi^2 \theta^t \Sigma[X] \theta} |e^{\delta(\theta)} - 1| d\theta,$$

where $|\delta(\theta)| = O(M\|\theta\|_R^3 N)$. Our assumption that $M\varepsilon^3 N \leq c$ for a sufficiently small c implies that $|e^{\delta(\theta)} - 1| \leq 2M\|\theta\|_R^3 N$ for $\theta \in \mathcal{B}_R(\varepsilon)$. Thus

$$\begin{aligned} I_1 &\leq 2MN \int_{\mathcal{B}_R(\varepsilon)} e^{-2\pi^2 \theta^t \Sigma[X] \theta} \|\theta\|_R^3 d\theta \leq 2MN \int_{\mathbb{R}^A} e^{-2\pi^2 \theta^t \Sigma[X] \theta} \|\theta\|_R^3 d\theta = \\ &= \frac{2MN}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma)}} \mathbb{E} [\|Z\|_R^3] \leq \frac{2M}{(2\pi^2)^{3/2} \sqrt{N} (2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma)}} \mathbb{E} [\|G\|_2^3], \end{aligned}$$

where we used again the relations (55) and (57). Finally, observing that by Jensen's inequality, $\mathbb{E}[\|G\|_2^3] \leq (\mathbb{E}[\|G\|_2^4])^{3/4} = (3|A| + |A|(|A| - 1))^{3/4} \leq 4^{3/4}|A|^{3/2}$ and that $\frac{2 \cdot 4^{3/4}}{(2\pi^2)^{3/2}} \leq \frac{1}{2}$ finishes the proof. \square

We make the choice

$$\varepsilon := \sqrt{\frac{2|A| \log N}{N}} \quad (58)$$

and the assumption

$$N \geq C' \cdot c_2 c_3^2 |A|^6 \log(2c_3 |A|)^6 \quad (59)$$

for some universal constant $C' > 0$ chosen sufficiently large for the following calculations. It is simple to check that with these choices the assumption $\varepsilon < \frac{1}{2M}$ as well as all the assumptions in the items of Lemma 4.14 hold. Thus, (50), (53) and Lemma 4.14 imply

$$\begin{aligned} \left| \mathbb{P}[X = \lambda] - \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det \Sigma[X]}} e^{-\frac{1}{2}(\lambda - \mathbb{E}[X])^t \Sigma[X]^{-1}(\lambda - \mathbb{E}[X])} \right| &\leq \\ &\leq \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma[X])}} \left(\frac{M|A|^{3/2}}{2\sqrt{N}} + \frac{(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma[X])}}{\det(\mathcal{L}(\phi))} e^{-\varepsilon^2 N} + e^{-\frac{1}{4}\pi^2 \varepsilon^2 N} \right). \end{aligned}$$

To compare the middle summand in the right-hand side with the others we use the following crude bounds on $\det(\mathcal{L}(\phi))$ and $\det(\Sigma[X])$. First, $\det(\mathcal{L}(\phi)) \geq 1$ since $\mathcal{L}(\phi) \subset \mathbb{Z}^A$. Second, since

$$\Sigma[X]_{a,a'} = p(1-p) \sum_{b \in B} \phi(b)_a \phi(b)_{a'} \leq c_2^2 N,$$

by Hadamard's inequality, we deduce that $\det(\Sigma[X]) \leq c_2^{2|A|} N^{|A|} |A|^{|A|/2}$. Thus, (58) and (59) imply

$$\frac{(2\pi)^{\frac{|A|}{2}} \sqrt{\det(\Sigma[X])}}{\det(\mathcal{L}(\phi))} e^{-\varepsilon^2 N} \leq \frac{(2\pi c_2^2 N)^{\frac{|A|}{2}} |A|^{\frac{|A|}{4}}}{N^{2|A|}} \leq \frac{1}{4N^{|A|/2}}$$

if the constant C' in (59) is large enough. Since also $e^{-\frac{1}{4}\pi^2 \varepsilon^2 N} \leq \frac{1}{4N^{|A|/2}}$ we finally conclude that

$$\mathbb{P}[X = \lambda] = \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det \Sigma[X]}} \left(e^{-\frac{1}{2}(\lambda - \mathbb{E}[X])^t \Sigma[X]^{-1}(\lambda - \mathbb{E}[X])} + \delta \right) \quad (60)$$

where $|\delta| \leq \frac{M|A|^{3/2}}{\sqrt{N}}$. Recalling that λ is an arbitrary point in $\mathcal{L}(\phi)$ and $N = p|B|$, we see that we have proven Theorem 4.1 in the case $p \leq \frac{1}{2}$.

We now get rid of the assumption $p \leq \frac{1}{2}$. Fix $p \geq \frac{1}{2}$, let $N := p|B|$ and assume that

$$|B| - N \geq C' \cdot c_2 c_3^2 |A|^6 \log(2c_3 |A|)^6 \quad (61)$$

holds. Recall that $X = \sum_{b \in B} T_b \phi(b)$ with the $\{T_b\}$ independent, identically distributed and satisfying $\mathbb{P}[T_b = 1] = 1 - \mathbb{P}[T_b = 0] = p$. Let us temporarily write $\mathbb{P}_p, \mathbb{E}_p$ and $\Sigma_p[X]$ for the probability, expectation and covariance matrix of X with a given p . Denote $\phi(B) := \sum_{b \in B} \phi(b)$. The fact that $X = \lambda$ if and only $\sum_{b \in B} (1 - T_b) \phi(b) = \phi(B) - \lambda$ implies that for any $\lambda \in \mathcal{L}(\phi)$, by (60), we have

$$\begin{aligned} \mathbb{P}_p[X = \lambda] &= \mathbb{P}_{1-p}[X = \phi(B) - \lambda] = \\ &= \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det \Sigma_{1-p}[X]}} \left(e^{-\frac{1}{2}(\phi(B) - \lambda - \mathbb{E}_{1-p}[X])^t \Sigma_{1-p}[X]^{-1}(\phi(B) - \lambda - \mathbb{E}_{1-p}[X])} + \delta \right) = \\ &= \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det \Sigma_p[X]}} \left(e^{-\frac{1}{2}(\lambda - \mathbb{E}_p[X])^t \Sigma_p[X]^{-1}(\lambda - \mathbb{E}_p[X])} + \delta \right), \end{aligned}$$

with $|\delta| \leq \frac{M|A|^{3/2}}{\sqrt{|B|-N}}$, as required. In the last equality we used the facts that $\Sigma_{1-p}[X] = \Sigma_p[X]$, $\mathbb{E}_{1-p}[X] = \phi(B) - \mathbb{E}_p[X]$ and that $\mu^t D \mu = (-\mu)^t D (-\mu)$ for any vector μ and matrix D . This establishes Theorem 4.1 in full.

4.7 Proof of main theorems

We now proceed to deduce Theorems 2.4 and 2.5 from the local central limit theorem, Theorem 4.1. Assume the conditions of Theorem 2.4 and let N be an integer satisfying condition (11). We wish to estimate the number of subsets $T \subset B$ of size N satisfying

$$\frac{1}{|T|} \sum_{t \in T} f(t) = \frac{1}{|B|} \sum_{b \in B} f(b) \quad \text{for all } f \text{ in } V. \quad (62)$$

Define $p := \frac{N}{|B|}$. Let $\{\phi_a : B \rightarrow \mathbb{Q}\}_{a \in A}$ be a basis for V . Define the random subset $T \subset B$ and random vector $X \in \mathbb{Q}^A$ as in Theorem 4.1. Since

$$X = \sum_{t \in T} \phi(t)$$

the event $X = \mathbb{E}[X]$ means

$$\sum_{t \in T} \phi(t) = p \sum_{b \in B} \phi(b) = \frac{N}{|B|} \sum_{b \in B} \phi(b).$$

Thus, since $\{\phi_a\}$, $a \in A$, is a basis for V , the event $X = \mathbb{E}[X]$ is equivalent to

$$\sum_{t \in T} f(t) = \frac{N}{|B|} \sum_{b \in B} f(b) \quad \text{for all } f \text{ in } V. \quad (63)$$

Now, by assumption, the constant function $h \equiv 1$ belongs to V . Thus, on the event $X = \mathbb{E}[X]$ we have

$$|T| = \sum_{t \in T} h(t) = \frac{N}{|B|} \sum_{b \in B} h(b) = N. \quad (64)$$

Comparing (62), (63) and (64) we see that the event $X = \mathbb{E}[X]$ is equivalent to the event that $|T| = N$ and (62) holds. Now, denoting by α_N the number of subsets $T \subset B$ of size N for which (62) holds it follows that

$$\mathbb{P}[X = \mathbb{E}[X]] = \alpha_N p^N (1-p)^{|B|-N}. \quad (65)$$

Finally, the divisibility assumption implies that

$$\mathbb{E}[X] = \frac{N}{|B|} \sum_{b \in B} \phi(b) \in \mathcal{L}(\phi).$$

Thus we may substitute $\lambda = \mathbb{E}[X]$ in Theorem 4.1 to obtain

$$\mathbb{P}[X = \mathbb{E}[X]] = \frac{\det(\mathcal{L}(\phi))}{(2\pi)^{\frac{|A|}{2}} \sqrt{\det \Sigma[X]}} (1 + \delta), \quad (66)$$

with $|\delta| \leq \frac{C \dim(V)^3 (\log(2c_2 \dim(V)))^{3/2}}{\sqrt{\min(N, |B|-N)}}$. Comparing (65) and (66) proves the assertion of Theorem 2.5.

Lastly, Theorem 2.4 follows upon noting that the assumption (11) implies that $|\delta| \leq \frac{1}{2}$, so that $\mathbb{P}[X = \mathbb{E}[X]] > 0$ and hence $\alpha_N > 0$.

4.8 Basis-free formulation of local central limit theorem

In this section we describe an equivalent “basis-free” version of our local central limit theorem, Theorem 4.1. The theorem is a high-dimensional, lattice, local central limit theorem with a rate of convergence estimate involving only universal constants.

Recall the parameter $\rho(V)$ of the vector space V introduced in (13). We introduce a second parameter of V , a non-negative definite form $\langle \cdot, \cdot \rangle_V$ on \mathbb{Q}^B . As is the case for $\rho(V)$, it is easiest to define $\langle \cdot, \cdot \rangle_V$ via a choice of basis for V but we stress that it is independent of this choice. If $\phi : B \rightarrow \mathbb{Q}^A$ is such that the vectors (ϕ_a) , $a \in A$, form a basis for V , we define

$$\langle \gamma_1, \gamma_2 \rangle_V := \gamma_1^t \phi (\phi^t \phi)^{-1} \phi^t \gamma_2 \quad (\gamma_1, \gamma_2 \in \mathbb{Q}^B).$$

In this definition, ϕ is regarded as a $B \times A$ matrix with columns $\{\phi_a\}$. The matrix $\phi (\phi^t \phi)^{-1} \phi^t$ represents the orthogonal projection operator from \mathbb{Q}^B (with the standard basis and inner product) to V . We denote the semi-norm induced from $\langle \cdot, \cdot \rangle_V$ by $\| \cdot \|_V$,

$$\| \gamma \|_V := \sqrt{\langle \gamma, \gamma \rangle_V} \quad (\gamma \in \mathbb{Q}^B),$$

so that $\| \gamma \|_V$ is the length of the orthogonal projection of γ to V . Finally, we denote by $\mathbf{1}$ the identically one vector in \mathbb{Q}^B .

Theorem 4.15 (Basis-free formulation of local central limit theorem). *There exists a constant $C > 0$ such that the following is true. Let B be a finite set and let V be a linear subspace of functions $f : B \rightarrow \mathbb{Q}$. Assume that the following conditions hold for some integers $c_2, c_3 \geq 1$,*

1. *Boundedness of V : V has a c_2 -bounded integer basis in ℓ_∞ .*
2. *Boundedness of V^\perp : V^\perp has a c_3 -bounded integer basis in ℓ_1 .*
3. *Symmetry: for any $b_1, b_2 \in B$ there exists a symmetry π of V satisfying $\pi(b_1) = b_2$.*

Let $0 < p < 1$ and form a random subset $T \subset B$ by taking each element of B into T independently with probability p . If

$$\min(p|B|, (1-p)|B|) \geq C \cdot c_2 c_3^2 \dim(V)^6 \log(2c_3 \dim(V))^6$$

then for every $\gamma \in \mathbb{Z}^B$ the probability of the event

$$\sum_{t \in T} f(t) = \sum_{b \in B} \gamma_b f(b) \quad \text{for all } f \text{ in } V \tag{67}$$

equals

$$\frac{\rho(V)}{(2\pi p(1-p))^{\frac{\dim(V)}{2}}} \left(\exp \left(- \frac{\| \gamma - p \cdot \mathbf{1} \|_V^2}{2p(1-p)} \right) + \delta(\gamma) \right) \tag{68}$$

$$\text{with } |\delta(\gamma)| \leq \frac{C \dim(V)^3 (\log(2c_2 \dim(V)))^{3/2}}{\sqrt{\min(p|B|, (1-p)|B|)}}.$$

It is important to emphasize that one must take $\gamma \in \mathbb{Z}^B$, rather than $\gamma \in \mathbb{Q}^B$ in the theorem, analogously to the restriction that $\lambda \in \mathcal{L}(\phi)$ in Theorem 4.1. However, since V typically has dimension strictly less than $|B|$, it is possible that for some $\gamma \in \mathbb{Q}^B \setminus \mathbb{Z}^B$ there exists another $\gamma' \in \mathbb{Z}^B$ such that

$$\sum_{b \in B} \gamma_b f(b) = \sum_{b \in B} \gamma'_b f(b) \quad \text{for all } f \text{ in } V. \tag{69}$$

Indeed, this is exactly the scenario we face in our main theorems. There, we are interested in the case that $\gamma_b = \frac{N}{|B|}$ for all $b \in B$, a vector which is not in \mathbb{Z}^B . The divisibility condition in Theorem 2.4 exactly ensures that for this vector there exists some $\gamma' \in \mathbb{Z}^B$ such that (69) holds.

Let us say that a vector $\gamma \in \mathbb{Q}^B$ has an *integer representation* by $\gamma' \in \mathbb{Z}^B$ if (69) holds. It is not difficult to check that in this case $\|\gamma - p \cdot \mathbf{1}\|_V = \|\gamma' - p \cdot \mathbf{1}\|_V$. Thus, our theorem remains true as stated if the restriction that $\gamma \in \mathbb{Z}^B$ is replaced by the condition that $\gamma \in \mathbb{Q}^B$ and has an integer representation. Moreover, it is evident that if γ has no integer representation then the probability of (67) is zero, since the left hand side of (67) exactly provides an integer representation for γ .

We now briefly explain the equivalence of Theorem 4.1 and Theorem 4.15. Suppose that $\{\phi_a\}$, $a \in A$, form a basis for V , define $\phi : B \rightarrow \mathbb{Q}^A$ by $\phi(b)_a = \phi_a(B)$ and regard ϕ as a $B \times A$ matrix. If $\lambda \in \mathcal{L}(\phi)$ then, by definition, there exists some $\gamma \in \mathbb{Z}^B$ such that

$$\lambda = \sum_{b \in B} \gamma_b \phi(b) = \phi^t \gamma. \quad (70)$$

It is then straightforward to check that for a subset $T \subset B$ condition (67) is equivalent to

$$\sum_{t \in T} \phi(t) = \lambda. \quad (71)$$

Conversely, given $\gamma \in \mathbb{Z}^B$ we may define $\lambda \in \mathcal{L}(\phi)$ by (70) and observe again that (67) is equivalent to (71). Thus, to see the equivalence of the two theorems, it suffices to show that the main terms in the probability estimates (36) and (68) are equal when the relation (70) holds. This follows from the definitions of $\rho(V)$, the definitions of $\mathbb{E}[X]$ and $\Sigma[X]$ in Subsection 4.1 and the observation that under (70) we have

$$\begin{aligned} \|\gamma - p \cdot \mathbf{1}\|_V^2 &= (\gamma - p \cdot \mathbf{1})^t \phi(\phi^t \phi)^{-1} \phi^t (\gamma - p \cdot \mathbf{1}) = \\ &= (\lambda - \mathbb{E}[X])^t (\phi^t \phi)^{-1} (\lambda - \mathbb{E}[X]) = p(1-p)(\lambda - \mathbb{E}[X])^t \Sigma[X]^{-1} (\lambda - \mathbb{E}[X]). \end{aligned}$$

5 Summary and open problems

Our main theorem guarantees the existence of a small subset $T \subset B$ for which (1) holds. The conditions we require are boundedness, divisibility and symmetry. In many natural scenarios it is easy to guarantee that V has a bounded integer basis in ℓ_∞ , the divisibility and the symmetry condition, and the condition which seems hardest to verify is that V^\perp has a bounded integer basis in ℓ_1 . In particular, the following question captures much of the difficulty. Let G be a group that acts transitively on a set X . A subset $T \subset G$ is *X-uniform* (or an *X-design*) if it acts on X exactly as G does. That is, for any $x, y \in X$,

$$\frac{1}{|T|} |\{g \in T : g(x) = y\}| = \frac{1}{|G|} |\{g \in G : g(x) = y\}| = \frac{1}{|X|}.$$

In our language we may take $B = G$ and V to be the space spanned by all functions $\phi_{(x,y)} : B \rightarrow \{0, 1\}$ of the form $\phi_{(x,y)}(b) = \mathbf{1}_{\{b(x)=y\}}$ for $x, y \in X$. Then T is X -uniform if and only if (1) holds. We have given a bounded integer basis for V in ℓ_∞ , and also by definition the symmetry condition holds. The other conditions are less clear. One may still speculate that:

Conjecture 5.1. *Let G be a group that acts transitively on a set X . Then there exists an X -uniform subset $T \subset G$ such that $|T| \leq |X|^c$ for some universal constant $c > 0$.*

A second question is whether one can apply our techniques to get *minimal* objects. Recall that the size of the objects we achieve is only minimal up to polynomial factors. For example, one of the main open problems in design theory is whether there exists a Steiner system (i.e. a t -design with $\lambda = 1$) for any $t > 5$. Another major open problem of a similar spirit is the existence of Hadamard matrices of all orders $n = 4m$, or equivalently, $2-(4m-1, 2m-1, m-1)$ designs. Empirical estimates for $n \leq 32$ suggest that there are $\exp(O(n \log n))$ Hadamard matrices of order $n = 4m$. Since there are so many of them, and since the logarithm of their number grows at a regular rate, we suspect that they exist for some purely statistical reason. However, the Gaussian local limit model seems to be false for Hadamard matrices interpreted as t -designs; it does not accurately estimate how many there are.

A third question is whether there exists an algorithmic version of our work, similar to the algorithmic Moser [Mos09] and Moser-Tardos [MT10] versions of the Lovász local lemma [EL75], and the algorithmic Bansal [Ban10] and Lovett-Meka [LM12] versions of the six standard deviations method of Spencer [Spe85]. If an efficient randomized algorithm of our method were found, then we could no longer indisputably claim that we have a low-probability version of the probabilistic method. On the other hand it would be strange, from the viewpoint of computational complexity theory, if low-probability existence can always be converted to high-probability existence. Maybe our construction is fundamentally a low-probability construction.

It is also of interest to extend our results to continuous setups, with one representative example being that of spherical designs (see, e.g., [BRV10] and references within). We aim to develop this direction in future work.

Acknowledgements. We thank David Soudry and Gady Kozma for useful remarks on the representation theory of the symmetric group.

References

- [AL11] Noga Alon and Shachar Lovett, *Almost k -wise vs k -wise independent permutations, and uniformity for general group actions*, 2011, ECCC TR11-049.
- [AV97] Noga Alon and Van H. Vu, *Anti-Hadamard matrices, coin weighing, threshold gates and indecomposable hypergraphs*, J. Combin. Theory Ser. A **79** (1997), no. 1, 133–160.
- [Ban10] Nikhil Bansal, *Constructive algorithms for discrepancy minimization*, Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, FOCS '10, IEEE Computer Society, 2010, arXiv:1002.2259, pp. 3–10.
- [Bap00] R. B. Bapat, *Moore-Penrose inverse of set inclusion matrices*, Linear Algebra Appl. **318** (2000), no. 1-3, 35–44.
- [BRV10] Andriy Bondarenko, Danylo Radchenko, and Maryna Viazovska, *Optimal asymptotic bounds for spherical designs*, arxiv:1009.4407.
- [Cam95] P. J. Cameron, *Permutation groups*, Handbook of combinatorics, Vol. 1, 2, Elsevier, 1995, pp. 611–645.
- [CD07] Charles J. Colbourn and Jeffrey H. Dinitz (eds.), *The CRC handbook of combinatorial designs*, 2nd ed., Discrete Mathematics and its Applications, Chapman & Hall/CRC, 2007.

- [CM05] E. Rodney Canfield and Brendan D. McKay, *Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums*, Electron. J. Combin. **12** (2005), Research Paper 29, 31 pp. (electronic).
- [EL75] Paul Erdős and László Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and Finite Sets, Coll. Math. Soc. J. Bolyai, no. 11, North-Holland, 1975, pp. 609–627.
- [Far09] B. Farhi, *An identity involving the least common multiple of binomial coefficients and its application*, Amer. Math. Monthly **116** (2009), no. 9, 836–839.
- [FH91] W. Fulton and J. Harris, *Representation theory*, Graduate Texts in Mathematics, vol. 129, Springer, New York, 1991.
- [FPY12] Hilary Finucane, Ron Peled, and Yariv Yaari, *A recursive construction of t -wise uniform permutations*, arxiv:1111.0492.
- [GJ73] J. E. Graver and W. B. Jurkat, *The module structure of integral designs*, J. Combinatorial Theory Ser. A **15** (1973), 75–90.
- [HSS99] A. S. Hedayat, N. J. A. Sloane, and John Stufken, *Orthogonal arrays: Theory and applications*, Springer-Verlag, 1999.
- [Jam78] G. D. James, *The representation theory of the symmetric groups*, Lecture Notes in Mathematics, vol. 682, Springer, Berlin, 1978.
- [KLP12] Greg Kuperberg, Shachar Lovett, and Ron Peled, *Probabilistic existence of rigid combinatorial structures*, Proceedings of the 44th annual ACM symposium on Theory of computing, STOC, ACM, 2012, arXiv:1111.0492, pp. 1091–1106.
- [KM94] Daphne Koller and Nimrod Megiddo, *Constructing small sample spaces satisfying given constants*, SIAM J. Discrete Math. **7** (1994), no. 2, 260–274.
- [KNR05] E. Kaplan, M. Naor, and O. Reingold, *Derandomized constructions of k -wise (almost) independent permutations*, Approximation, randomization and combinatorial optimization (C. Chekuri, K. Jansen, J. D. P. Rolim, and L. Trevisan, eds.), Lecture Notes in Computer Science, vol. 3624, Springer, 2005, pp. 354–365.
- [KP82] Richard M. Karp and Christos H. Papadimitriou, *On linear characterizations of combinatorial optimization problems*, SIAM J. Comput. **11** (1982), no. 4, 620–632.
- [LM12] Shachar Lovett and Raghu Meka, *Constructive discrepancy minimization by walking on the edges*, 2012, arXiv:1203.5747.
- [Mag09] Spyros S. Magliveras, *Large sets of t -designs from groups*, Mathematica Slovaca **59** (2009), no. 1, 1–20.
- [Mos09] Robin A. Moser, *A constructive proof of the Lovász local lemma*, Proceedings of the 41st annual ACM symposium on Theory of computing, STOC, ACM, 2009, arXiv:0810.4812, pp. 343–350.
- [MT10] Robin A. Moser and Gábor Tardos, *A constructive proof of the general Lovász local lemma*, J. ACM **57** (2010), no. 2, 11:1–11:15, arXiv:0903.0544.

- [MW90] Brendan D. McKay and Nicholas C. Wormald, *Asymptotic enumeration by degree sequence of graphs of high degree*, European J. Combin. **11** (1990), no. 6, 565–580. MR 1078713 (91j:05010)
- [Rao73] C. Radhakrishna Rao, *Some combinatorial problems of arrays and applications to design of experiments*, Survey of combinatorial theory (J. N. Srivastava, ed.), North-Holland, 1973, pp. 349–359.
- [RCW75] Dijen K. Ray-Chaudhuri and Richard M. Wilson, *On t -designs*, Osaka J. Math. **12** (1975), no. 3, 737–744.
- [Rob38] G. de B. Robinson, *On the representations of the symmetric group*, American Journal of Mathematics **60** (1938), no. 3, 745–760.
- [Sch61] C. Schensted, *Longest increasing and decreasing subsequences*, Canadian Journal of Mathematics **13** (1961), 179–191.
- [Spe85] Joel Spencer, *Six standard deviations suffice*, Trans. Amer. Math. Soc. **289** (1985), no. 2, 679–706.
- [SZ84] P. D. Seymour and Thomas Zaslavsky, *Averaging sets: a generalization of mean values and spherical designs*, Adv. in Math. **52** (1984), no. 3, 213–240.
- [Tei87] Luc Teirlinck, *Non-trivial t -designs without repeated blocks exist for all t* , Discrete Math. **65** (1987), no. 3, 301–311.
- [Wil73] Richard M. Wilson, *The necessary conditions for t -designs are sufficient for something*, Utilitas Math. **4** (1973), 207–215.
- [Wil90] ———, *A diagonal form for the incidence matrices of t -subsets vs. k -subsets*, European J. Combin. **11** (1990), no. 6, 609–615.
- [Yek11] Sergey Yekhanin, *Locally decodeable codes*, Foundations and Trends in Theoretical Computer Science **7** (2011), no. 1, 1–117.